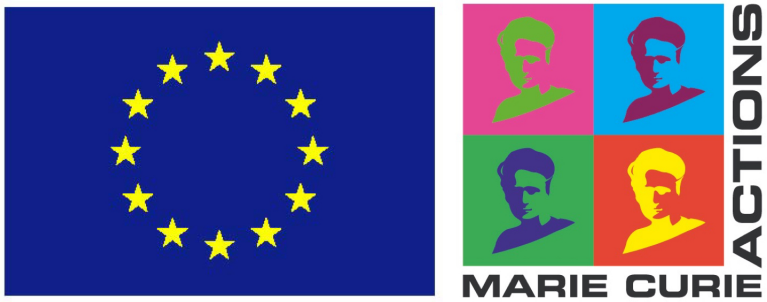
Algoritmi equi ed educazione civica

Materiale didattico

Michele Loi

Versione Beta
31 maggio 2023

Progetto finanziato dal programma di ricerca e innovazione Horizon 2020 dell'Unione Europea nell'ambito dell'accordo di sovvenzione Marie Sklodowska-Curie n. 89832, "Fair Predictions in Health" e svolto in collaborazine con il laboratorio FDS (Formazione e Sperimentazione Didattica) del Politecnico di Milano.



Introduzione

Al giorno d'oggi, chi progetta e utilizza strumenti basati sull'intelligenza artificiale, non può fare a meno di considerare le ricadute etiche e sociali delle decisioni prese nello svolgere questo ruolo.

Per capire i problemi della discriminazione e della iniquità algoritmica è necessario considerare contemporaneamente sia gli aspetti matematici sia gli aspetti quelli morali.

Queste schede sono state ideate per gli insegnanti di matematica, filosofia, diritto ed economia con lo scopo di progettare e realizzare un percorso di educazione civica interdisciplinare sul tema della discriminazine e della equità algoritmica.

Alla fine di questo itinerario didattico, gli studenti saranno in grado di orientarsi attraverso il mondo dell'equità algoritmica sia dal punto di vista matematico sia dal punto di vista filosofico. Saranno in grado di comprendere il genere di domande che occorre porsi per capire quando un algoritmo discrimina o produce risultati iniqui. D'altra parte gli studenti potranno anche formulare risposte moralmente e matematicamente plausibili su come ridurre l'iniquità. A tal fine, l'itinerario didattico si sviluppa attraverso le seguenti dieci schede.

È importante sottolineare che questo lavoro è allo stadio attuale un lavoro di ricerca, che necessita di una fase di sperimentazione utile a selezionare e affinare le attività didattiche che i dicenti possono sviluppare a partire dalle tematiche trattate[1].

Scheda 1. La discriminazione indiretta e la diseguaglianza tra gruppi.

Scheda 2. Introduzione alle decisioni algoritmiche.

Scheda 3. Regole soglia ed effetti sociali.

Scheda 4. Diseguaglianza algoritmica tra gruppi.

[1] Eventuali osservazioni possono essere inviate a michele.loi@polimi.it.

1. La discriminazione indiretta e la diseguaglianza tra gruppi

Le linee rosse di Amazon

Nel 2016, la diffusione del servizio di Amazon "Consegna gratuita in un giorno" in molte città degli Stati Uniti mostrava una corrispondenza molto evidente con la distribuzione delle diverse etnie (o *race* [razza] nell'accezione nordamericana del termine)[2] all'interno dei diversi quartieri di quelle città. Questo accade nonostante il fatto che l'algoritmo utilizzato per decidere a quali parti della città estendere il servizio non abbia mai usato i dati sulla appartenenza dei suoi consumatori a gruppi etnici o *race group*[3]. Ciò che fa l'algoritmo è semplicemente valutare se fornire il servizio aumenterebbe i profitti, in base ai pattern di acquisti dei consumatori delle diverse zone. Nonostante ciò, dopo la pubblicazione dell'articolo, l'esito è stato ritenuto inaccettabile dai politici e Amazon ha subito modificato le sue precedenti scelte estendendo il servizio a tutti i quartieri, anche quelli fino a quel momento esclusi. A prescindere dalla intenzione e dalle ragioni dell'azienda, la pratica risultante ricorda il cosiddetto fenomeno delle linee rosse di banche e assicurazioni che usavano il codice di avviamento postale dei clienti per calcolare il rischio finanziario e questo finiva per escludere da tali servizi le zone abitate in prevalenza dalle minoranze. Queste pratiche sono successivamente divenute illegali.

[2]Per rigore sociologico, non è possibile evitare di usare il concetto di *race* in questo esempio, in quanto tale è il concetto sociologicamente rilevante per gli Stati Uniti. Nelle discipline sociologiche nordamericane il concetto può essere usato in riferimento al fenomeno sociologico senza alcuna pretesa di affermare la realtà della razza in senso naturalistico. Sarà compito degli insegnanti (es. di filosofia e di inglese) valutare come introdurre il problema dell'uso del concetto di "razza" nella critica e nella prassi anti-razzista, che è un tema di discussione contemporanea, e quello della traduzione del termine *race* in italiano, che richiede attenzione al contesto.

[3]Si veda la nota [2].

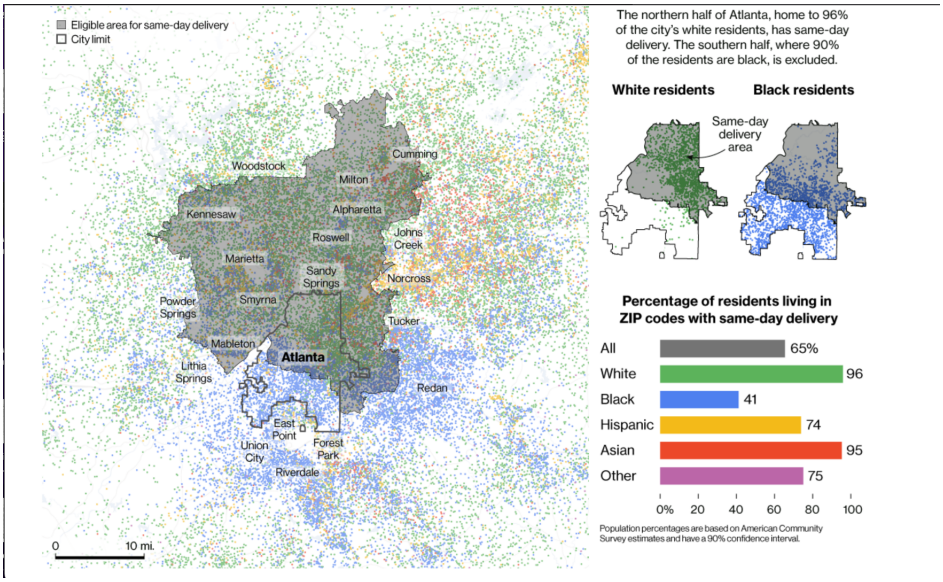


Figura 1: L'immagine rappresenta la mappa di Atlanta, in cui la metà settentrionale include 96% dei residenti bianchi ed è servita dal servizio *consegna grautita in un giorno*. La metà meridionale, in cui il 90% dei residenti è nero, è esclusa. L'istogramma rappresenta la distribuzione dei gruppi nelle zone in cui il servizio è presente. Come riportato nella versione aggiornata dell'articolo (link fornito in bibliografia), nel corso del tempo il servizio è stato esteso all'intera città. Immagine tratta da [14] e utilizzata per finalità illustrative in base all'Articolo 70 – Legge 633/1941 sul diritto d'autore - non è concesso l'utilizzo di questa immagine per finalità commerciali.

Il caso COMPAS

COMPAS, Correctional Offender Management Profiling for Alternative Sanctions[4] è un sistema di predizione commerciale (prodotto dalla Northponte) per predire il "recidivismo", termine utilizzato per indicare un criminale che compie un crimine dopo essere stato rilasciato dal carcere.

L'organizzazione giornalistica ProPublica [1] si è chiesta se le persone, che dopo essere uscite dalla prigione non avevano commesso alcun reato nei 2 anni successivi (chiamati perciò "non recidivisti") avessero una diversa probabilità di essere classificati come criminali ad alto rischio di recidivismo a seconda della loro appartenenza al gruppo dei bianchi, dei neri o degli ispanici. Hanno scoperto che la probabilità dei non recidivisti di essere considerati ad alto rischio era circa due volte più alta nel caso dei neri rispetto ai bianchi. Il caso è stato ampiamente discusso nella stampa e in accademia. Uno dei risultati del dibattito è stato capire che il

[4]Profilazione per la gestione correttiva dei criminali con pene alternative.

criterio utilizzato da ProPublica per valutare l'equità di COMPAS non è l'unico plausibile. In risposta alle critiche ricevute, l'azienda che ha sviluppato COMPAS ha mostrato che le predizioni risultavano egualmente accurate tra bianchi e neri, utilizzando una misura statistica *diversa* della accuratezza [5]. Grazie a questo caso, la comunità scientifica ha preso atto della complessità etica della questione su quale sia il criterio di accuratezza più sensato nel contesto di una decisione che riguarda gli esseri umani [7].

2. Introduzione alle decisioni algoritmiche

Che cos'è una "buona" decisione algoritmica? Da un certo punto di vista, essa è una decisione vantaggiosa per il decisore: quella che crea valore per il decisore. Osserviamo, ad esempio, la seguente tabella.

Decisore	**Tipo di decisione**	**Valore per il decisore (Utilità)**
Banca	Fornire un prestito	Ottenere profitto dagli interessi
Giudice	Rilasciare una detenuto per buona condotta	Reinserimento in società senza recidività (si veda Scheda 1)
Azienda	Assumere il candidato per una posizione lavorativa	Realizzazione degli obiettivi legati al profilo lavorativo

Schema dei processi decisionali

Consideriamo il caso della decisione binaria (sì/no). Nel caso del prestito bancario possiamo sintetizzare il processo grazie al seguente schema nel qual introduciamo le variabili D e Y che indicano rispettivamente la decisione e la variabile rilevante per prendere la decisione.

	Y = non ripaga il prestito	Y = ripaga il prestito
D = no (non concedo prestito)	0€	0€
D = sì (concedo il prestito)	-1000€	100€

Tabella 1: Tabella di utilità

Il problema di scegliere una soglia di probabilità adeguata è quello che in matematica viene chiamato problema di ottimizzazione. L'interesse del decisore è quello di massimizzare l'utilità. Chiaramente, è possibile stabilire più regole di decisione che portano a utilità differenti. La questione diventa interessante quando il decisore non ha la piena conoscenza della variabile rilevante Y. In tal caso, subentra il concetto di incertezza e quello ad esso strettamente collegato di probabilità.
Per comprendere meglio la dinamica del processo decisionale affetto da incertezza consideriamo l'esempio precedente nel quale la variabile Y assume il valore "ripaga il prestito" con probabilità 0.98 e il valore "non ripaga il prestito" con probabilità 0.02. A questo punto, entra in gioco la nozione di valore atteso di una variabile (in questo caso, l'utilità del decisore U), che possiamo definire nel seguente:

$$\mathbb{E}(U) = \mathbb{P}(Y = \text{ripaga il prestito}) \cdot (\text{utilità se ripaga il prestito}) +$$
$$\mathbb{P}(Y = \text{non ripaga il prestito}) \cdot (\text{utilità (negativa) se non ripaga il prestito})$$

che in questo esempio equivale a

$$\mathbb{E}(U) = 0.02 \cdot (-1000) + 0.98 \cdot (+100) = -20 + 98 = 78$$

per $D =$ sì (concedo il prestito), mentre equivale a

$$E(U) = 0.02 \cdot (0) + 0.98 \cdot (0) = 0 + 0 = 0$$

per $D =$ no (non concedo prestito).

Risulta ora evidente che la scelta che massimizza l'utilità del decisore date le probabilità in gioco è la prima, cioè concedere il prestito.

Tuttavia, talvolta capita che la concessione del prestito non sia la scelta che massimizza l'utilità del decisore.

Ad esempio, se fissiamo $P(Y = \text{ripaga il prestito}) = 0.70$, abbiamo

$$\mathbb{E}(U) = 0.30 \cdot (-1000) + 0.70 \cdot (+100) = -300 + 70 = -230.$$

Notiamo ora che la riga relativa alla decisione $D =$ no (non concedo prestito) ci fornisce nuovamente un'utilità attesa pari a zero, che però questa volta risulta essere la decisione migliore:

$$\mathbb{E}(U) = 0.30 \cdot (0) + 0.70 \cdot (0) = 0.$$

3. Regole soglia ed effetti sociali

Quanto visto nella scheda precedente si può definire come il calcolo del valore atteso dell'utilità per il decisore che valuta un singolo individuo. Realisticamente le banche compiono questo tipo operazioni per tutti gli individui che richiedono un prestito. Emerge quindi la necessità, da parte del decisore, di stabilire una regola di decisione generale che tenga conto dell'incertezza. Come procedere?

Si può dimostrare matematicamente che la decisione ottimale D in funzione della probabilità p è data dalla seguente regola (denominata "regola soglia"):

$$D(p) = \begin{cases} \text{sì} & \text{se } p \geq p_0 \\ \text{no} & \text{altrimenti} \end{cases}$$

dove p_0 deve essere calcolato tenendo conto di tutti i valori della tabella di utilità.

Esempio

Consideriamo il caso studiato nella Scheda 2 relativo al rilascio di un prestito presso un istituto bancario.

	Y = non ripaga il prestito	Y = ripaga il prestito
D = no (non concedo prestito)	0€	0€
D = sì (concedo il prestito)	-1000€	100€

Esaminiamo la seguente regola soglia definita dalle seguenti probabilità:

$$p_0 = 0.98 \quad \text{e} \quad (1 - p_0) = 0.02.$$

Per ogni $p \geq p_0$, per la riga relativa alla decisione D = sì (concedo il prestito) abbiamo (sempre):

$$\mathbb{E}(U) = (1 - p) \cdot (-1000) + p \cdot (+100) \geq 0,$$

e per la riga relativa alla decisione D = no (non concedo prestito) abbiamo (sempre)

$$\mathbb{E}(U) = (1-p)\cdot(0) + p\cdot(0) = 0.$$

Esercizio esplicativo: come costruire la tabella di decisione per un gruppo sociale

Consideriamo il caso studio definito dalla seguente tabella in cui il valore soglia p_0 è fissato a 0.70. Per comodità, quando gli esiti possibili sono due (il cosiddetto caso *binario*) spesso si utilizza la notazione 0 ed 1 per indicare gli assegnamenti rispettivamente negativi e positivi alle variabili D ed Y. Per **Predizione** si intende la probabilità di ripagare il debito stimata per ciascun individuo in base alle sue caratterestiche individuali note alla banca.

ID	**Predizione**	**Decisione** D	**Sesso**	**Variabile Rilevante** Y
1	0.81	1	m	0
2	0.80	1	m	1
3	0.79	1	m	1
4	0.77	1	m	1
5	0.75	1	f	1
6	0.74	1	m	1
7	0.74	1	m	1
8	0.73	1	m	1
9	0.72	1	f	0
10	0.72	1	f	1
11	0.68	0	m	1
12	0.68	0	m	1
13	0.68	0	f	1
14	0.62	0	f	0
15	0.60	0	f	1
16	0.55	0	m	0
17	0.48	0	f	1
18	0.47	0	f	0
19	0.40	0	f	1
20	0.37	0	f	0

Tabella 2: Database dei richiedenti prestito

Possiamo ora completare la tabella 3, nella pagina seguente, che prende il nome di **tabella di decisione**. È importante sottolineare che la tabella 3 è simile

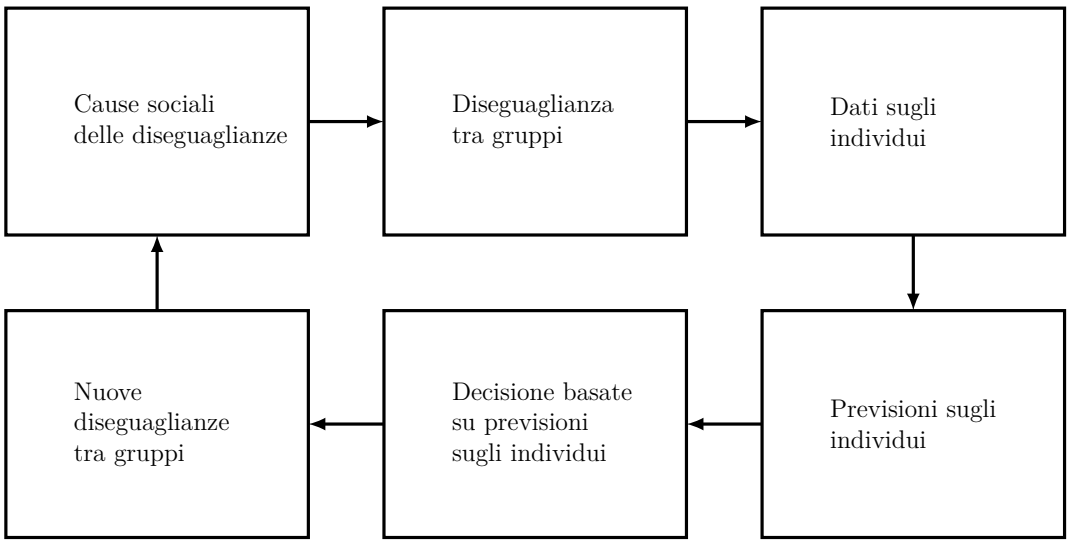
a quella vista in precedenza, ma ha un significato profondamente differente. Infatti, essa evidenzia la distribuzione assoluta per uno specifico gruppo sociale (ad esempio, gli "uomini", m) della decisione D rispetto alla variabile rilevante Y.

	$Y = 0$: non ripaga il prestito	$Y = 1$: ripaga il prestito
$D = 0$: no (non concedo prestito)	1	2
$D = 1$: sì (concedo il prestito)	1	6

Tabella 3: Tabella di decisione per il gruppo "uomini"

Cosa succede se costruiamo la medesima tabella per il gruppo "donne" (f)? Cosa possiamo dedurre, in particolare dalla casella individuata dalla decisione $D =$ sì (concedo il prestito) e $Y =$ ripaga il prestito?

4. Diseguaglianza algoritmica tra gruppi



Nelle nostre società, alcune cause sociali come sessismo, razzismo o le minori opportunità per gli individui che nascono nei contesti socialmente ed economicamente più svantaggiati generano diseguaglianze (es., di reddito, o qualifiche lavorative) tra i gruppi, che non sono moralmente giustificate [13] (si veda la Scheda 5 per il concetto di diseguaglianze giustificate). Queste diseguaglianze si riflettono nei dati utilizzati dagli algoritmi per compiere previsioni e su cui si basano le decisioni (sia quelle umane influenzate dalle raccomandazioni algoritmiche, sia quelle automatiche). A loro volta, tali decisioni possono contribuire alle diseguaglianze di partenza che caratterizzano la società. Ma che cosa si intende per *gruppo*, in questo contesto?

Definizione 1 Un gruppo è un insieme di individui che hanno una (o più caratteristiche in comune).

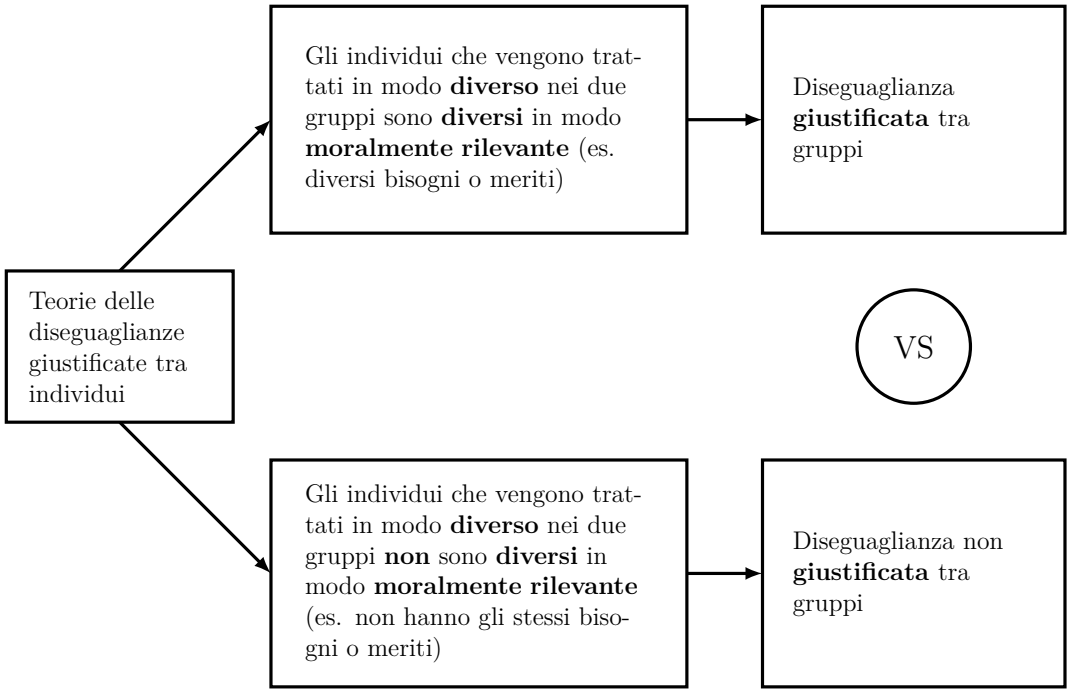
Definizione 2 Un gruppo è un insieme di individui che hanno una (o più) caratteristiche in comune che ha *salienza sociale*.

Che cosa si intende per *salienza sociale*?

Possiamo parlare di salienza sociale quando:

- appartenere al gruppo causa diseguaglianze (es. sessismo, razzismo, influenza della condizione socio-economico di partenza);
- il gruppo simboleggia uno status superiore / inferiore
- l'appartenenza al gruppo influisce in modo positivo nelle interazioni con altre persone dello stesso gruppo (ad esempio tramite la solidarietà).

Che cosa si intende per *diseguaglianze giustificate* tra gruppi?



Le diseguaglianze giustificate tra gruppi sono quindi quelle diseguaglianze che emergono come risultato delle diseguaglianze giustificate tra gli individui di cui tali gruppi sono formati. Il tema di cosa giustifichi moralmente le diseguaglianze tra individui è un tema classico della filosofia morale discusso anche recentementemente nella filosofia politica contemporanea di matrice analitica.

5. Le diseguaglianze giustificate

La giustizia, secondo Aristotele (*Etica Nicomachea*) consiste nel trattare in modo simile casi simili.

Questa concezione di giustizia si contrappone alla teoria della *uguaglianza semplice* secondo cui la giustizia consiste nella distribuzione uguale di qualcosa (es., le risorse, o il benessere individuale).

Contro l'eguaglianza semplice, possiamo menzionare teorie della giustizia in base alle quali la giustizia non consiste nell'eguaglianza semplice, ma nell'eguaglianza tra coloro che soddisfano *determinate condizioni moralmente rilevanti*.

Teorie della giustizia basate sui bisogni

Secondo questa classe di teorie, una distribuzione giusta fornisce uguali risorse a individui che hanno gli stessi bisogni di quelle risorse. Questa teoria risulta plausibile specialmente quando i bisogni in questione sono quelli sanitari e le risorse in questione sono le cure [20].

Teoria della giustizia basate sulla responsabilità

Secondo questa classe di teorie, una distribuzione giusta fornisce uguali vantaggi (o svantaggi) a persone che hanno volontariamente scelto esiti diversi (ad esempio, nel bilanciamento tra lavoro e svago), in virtù di scelte di cui possono essere considerati individualmente responsabili. Questa teoria risulta plausibile specialmente quando si considera la giustificazione delle diseguaglianze tra persone con diverse preferenze tra consumo e risparmio, tra risparmio rischioso o prudente, o tra coloro che giocano d'azzardo o si astengono da ciò. In base a questo genere di teoria, la differenza negli esiti che risulta da una scelta è considerata giustificata e non contraria al senso di giustizia [8, 2, 18, 10, 11].

Teorie della giustizia basate sul contributo (meritocrazia)

Secondo questa classe di teorie, una distribuzione giusta degli esiti di una attività cooperativa è sensibile al diverso grado di contributo di ciascuno a tale attività. Ad esempio, i lavoratori che contribuiscono maggiormente ai risultati di un'azienda dovrebbero ricevere salari più alti [15, 6].

6. Le critiche alla eguaglianza e il livellamento verso il basso

L'obiezione del livellamento verso il basso (Parfit) [16]. Immaginiamo che tutte le persone che nascono hanno un gemello privo di entrambi gli occhi. Se fosse possibile trapiantare un occhio, permettendo a entrambi di vedere, l'esito sarebbe migliore dal punto di vista della giustizia. O almeno, chi sostiene l'egalitarismo potrebbe plausibilmente sostenere questa tesi. È migliore perché è più eguale. Ma supponiamo che i due gemelli nascano uno con un occhio e l'altro senza. Se quello che conta moralmente è semplicemente l'eguaglianza, allora anche l'esito che si ottiene privando uno dei due gemelli del solo occhio che ha dovrebbe essere considerato migliore dal punto di vista della giustizia. "Questo esito sarebbe in qualche maniera migliore anche se non è in alcun modo meglio per il cieco. Troviamo questa proposizione impossibile da credere." ([16], trad. dell'autore)

Nelle tabelle qui sotto, alcuni esempi che possono essere utilizzati per illustrare il problema del livellamento verso il basso. Il terzo esempio riguarda le decisioni algoritmiche.

Decisioni politiche

Ricchezza		
	Ricchi	Poveri
Capitalismo (con welfare state)	10000€	2000€
Socialismo (reale)	1000€	1000€

Divisione delle torte

Dimensione della fetta		
	Io	Tu
Torta grande	400g	200g
Torta piccola	100g	100g

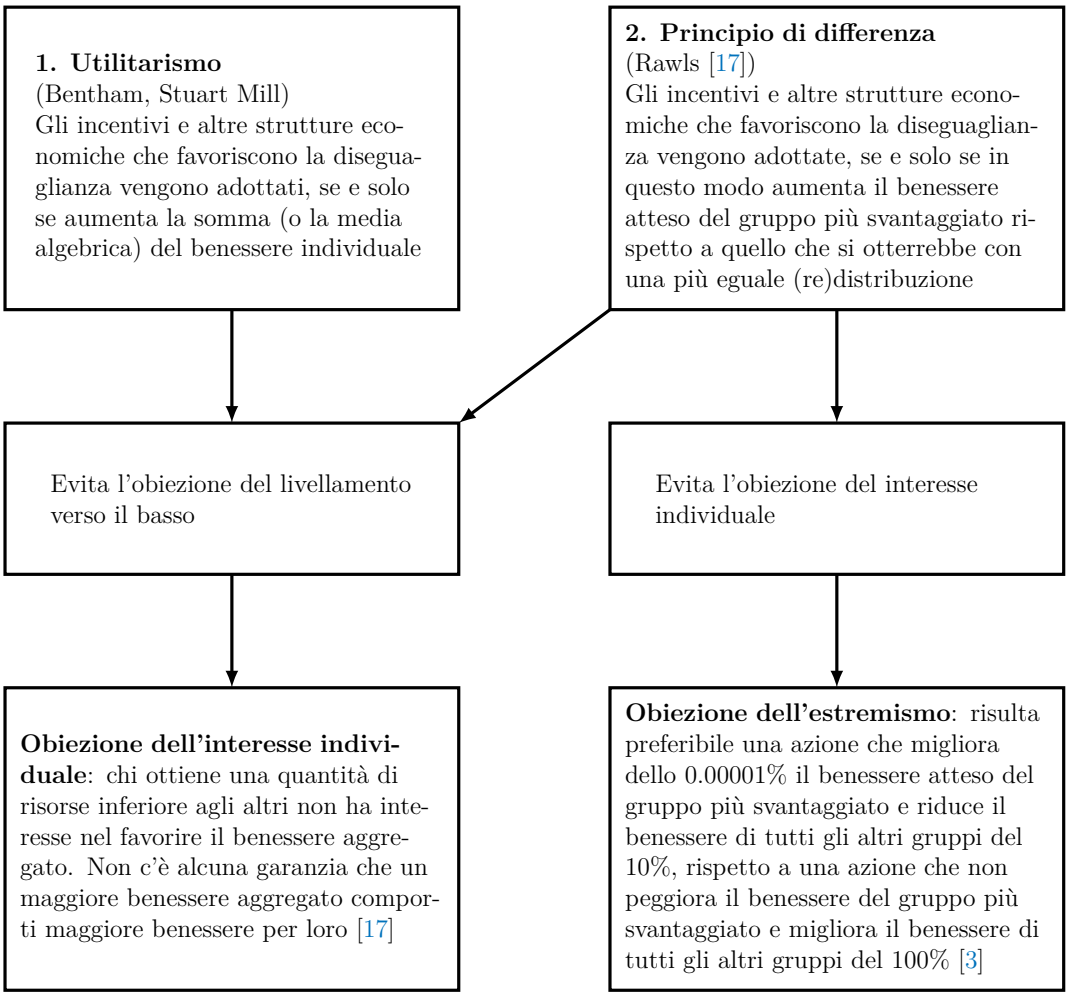
Decisioni algoritmiche

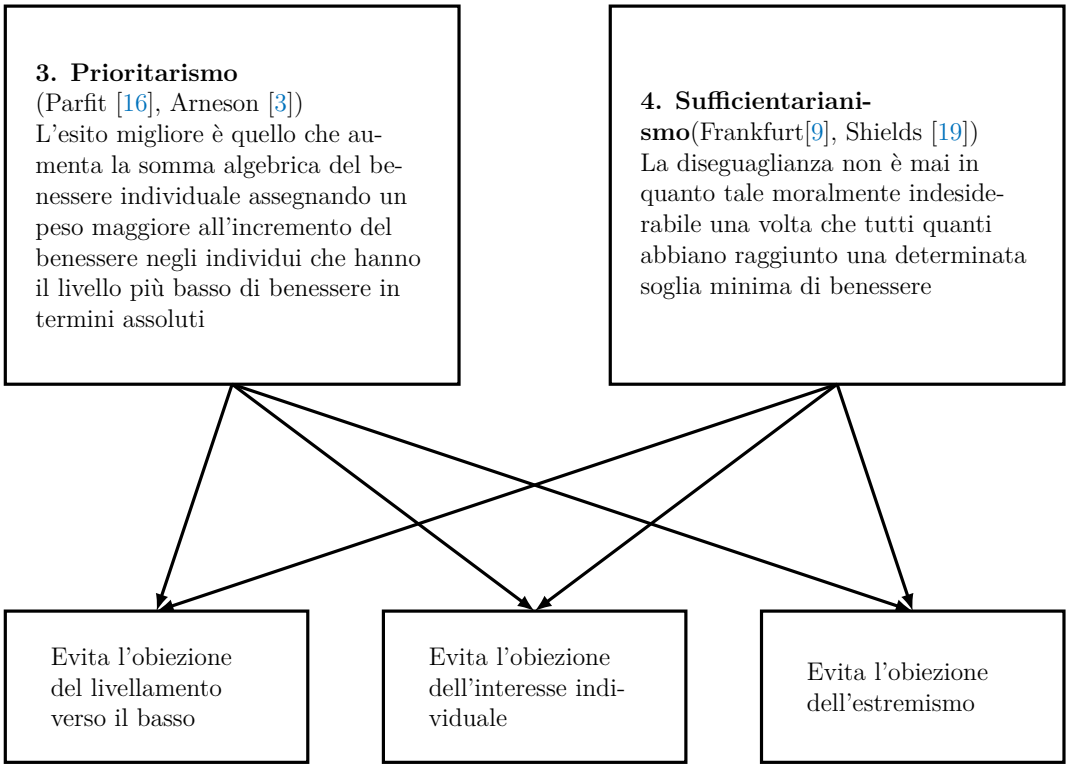
Predizioni corrette		
	Italiano	Immigrato
Algoritmo 1	80%	60%
Algoritmo 2	60%	60%

In tutti i tre casi, quando l'eguaglianza può essere ottenuta soltanto in modi che non comportano alcun vantaggio per la parte più svantaggiata l'argomento morale

per l'eguaglianza risulta molto indebolito. In altre parole, preferire l'eguaglianza nei tre casi, come nell'esempio di Parfit, espone alla critica del livellamento verso il basso.

7. Teorie della giustizia (alternative rispetto all'egalitarismo semplice)




3. Prioritarismo
(Parfit [16], Arneson [3])
L'esito migliore è quello che aumenta la somma algebrica del benessere individuale assegnando un peso maggiore all'incremento del benessere negli individui che hanno il livello più basso di benessere in termini assoluti
4. Sufficientarianismo(Frankfurt[9], Shields [19])
La diseguaglianza non è mai in quanto tale moralmente indesiderabile una volta che tutti quanti abbiano raggiunto una determinata soglia minima di benessere
Evita l'obiezione del livellamento verso il basso
Evita l'obiezione dell'interesse individuale
Evita l'obiezione dell'estremismo

8. Giustizia tra gruppi e decisioni algoritmiche

La Scheda 2 spiega come è possibile calcolare l'utilità attesa di un decisore (ad esempio, la banca) in condizioni di incertezza sull'esito Y, sulla base della probabilità p per un individuo della variabile di interesse Y. La Scheda 3 invece mostra che possiamo stabilire che una determinata regola soglia avrà un ritorno positivo per il decisore. In questa scheda consideriamo l'impatto che la scelta di una specifica regola di decisione ha sulle persone (il cosiddetto "Soggetto a Decisione (SD)"). Un esempio classico è il cliente della banca che richiede un mutuo. Più precisamente, nelle prossime righe cercheremo di calcolare l'utilità media per un gruppo che deriva dalla scelta, da parte del decisore, di una determinata regola soglia.

Per prima cosa è necessario determinare un modello matematico del modo in cui una determinata decisione algoritmica (considerata come un evento certo) impatta sulla persona che la subisce. Il modello matematico è una semplificazione della realtà. Nel modello che proponiamo, la semplificazione consiste in questo: esprimiamo il vantaggio o lo svantaggio di una decisione algoritmica per il SD come funzione di due soli parametri: la scelta ricevuta (ad esempio, il prestito viene concesso, oppure negato) e il comportamento di chi subisce la decisione, il Soggetto alla Decisione, SD (in questo caso: il debitore ripaga il prestito, oppure no). Questa idea può essere rappresentata dalla seguente *tabella di utilità del Soggetto alla Decisione (SD)*:

Il significato di questa tabella è il seguente:

- Casella $[W_{1,1}]$: l'utilità per l'individuo che non riceve un prestito e che non sarebbe comunque riuscito a ripagare il prestito (qualora l'avesse ricevuto) è pari a 0, cioè l'individuo non è né danneggiato né beneficiato dalla decisione.
- Casella $[W_{1,2}]$: l'utilità per l'individuo che non riceve un prestito e che sarebbe riuscito a ripagare il prestito (qualora l'avesse ricevuto) è pari a 1, cioè l'individuo è danneggiato lievemente da questa decisione.

	Y = non ripaga il prestito	Y = ripaga il prestito
D = no (non concedo prestito)	0 (casella $[W_{1,1}]$)	-1 (casella $[W_{1,2}]$)
D = sì (concedo il prestito)	-5 (casella $[W_{2,1}]$)	10 (casella $[W_{2,2}]$)

Tabella 4: Tabella di utilità del Soggetto alla Decisione

- Casella $[W_{2,1}]$: l'utilità per l'individuo che riceve un prestito che poi non riesce a ripagare è pari a -5, cioè l'individuo è danneggiato.

- Casella $[W_{2,2}]$: l'utilità per l'individuo che riceve un prestito che poi riesce a ripagare è pari a +10, cioè l'individuo è avvantaggiato dall'accesso a risorse finanziarie nel momento in cui gli servono di più (ad esempio, per l'acquisto di una casa).

Si noti che è la misura della utilità del SD nei quattro diversi casi può essere controversa. Nel nostro approccio, questi valori di utilità vengono definiti da un accordo intersoggettivo di esperti, che consideri l'opinione delle persone coinvolte. I valori da noi attribuiti rappresentano la sintesi di un possibile ragionamento sull'impatto delle decisioni. Ad esempio, noi abbiamo assegnato un valore basso di utilità negativa nel caso $[W_{1,2}]$, perché modelliamo la situazione in cui chi non riceve un prestito da una banca potrebbe riceverlo da una banca diversa. Inoltre, nel caso W2,1 abbiamo ritenuto plausibile un valore negativo poiché chi si ritrova a godere di un vantaggio temporaneo per la disponibilità finanziaria, ma alla fine viene incluso nella lista dei debitori insolventi e non riesce più a ottenere alcun credito). Per permettere la discussione di questi parametri, abbiamo sviluppato una piattaforma online che chiede allo user di inserire questi valori come base per le successive analisi dei dati di equità (Scheda 9).

Come si vede, descrivendo ogni evento di decisione sulla base di due parametri (D e Y) che hanno, ciascuno, due soli valori possibili, otteniamo quattro casi in tutto.

Il secondo passo consiste definire l'utilità media di ciascun gruppo in base alla **frequenza** dei quattro tipi di casi considerati, in quel gruppo specifico. Gruppo qui può significare qualsiasi insieme di individui di cui ci interessa calcolare il valore medio. Ma dal punto di vista della equità noi siamo interessati ai gruppi socialmente salienti, definiti nel senso della Scheda 4. L'utilità media di ciascun gruppo è quindi data da

$$\mathbb{E}_{SD}(G) = [W1, 1] \cdot \text{freq. di } [W_{1,1}] + [W_{1,2}] \cdot \text{freq. di } [W_{1,2}]$$
$$+[W_{2,1}] \cdot \text{freq. di } [W_{2,1}] + [W_{2,2}] \cdot \text{freq. di } [W_{2,2}],$$

dove le frequenze sono basate sulla (Scheda 3). Ad esempio, per il gruppo $G = \{\text{uomini}\}$ la tabella di decisione include i seguenti valori:

	Y=0	Y=1
D=0	1	2
D=1	1	6

Possiamo ora calcolare:

- frequenza di $[W_{1,1}] = 1/10$;
- frequenza di $[W_{1,2}] = 2/10$;
- frequenza di $[W_{2,1}] = 1/10$;
- frequenza di $[W_{2,1}] = 6/10$.

Esercizio

Calcolare l'utilità media del gruppo uomini e, in riferimento all'esercizio della scheda Scheda 3, l'utilità media del gruppo donne ($G = \{\text{donne}\}$), usando la formula $\mathbb{E}_{SD}(G)$.

L'ultimo passaggio consiste nello stabilire il rapporto tra la regola di decisione scelta dal decisore e la frequenza dei quattro tipi di casi in questione, che sono la frequenza di $[W_{1,1}]$, frequenza di $[W_{1,2}]$, frequenza di $[W_{2,1}]$, frequenza di $[W_{2,1}]$. Come verificato nell'esercizio della scheda Scheda 3, al cambiare della regola di decisione (ad esempio, la soglia di rischio impostata dall'algoritmo della banca nell'esercizio della Scheda 3), cambiano i valori nella tabella di decisione dei diversi gruppi. Da ciò derivano utilità medie diverse per i diversi gruppi.

9. Modelli matematici delle concezioni di giustizia

Nella Scheda 6 sono state presentate alcune concezioni di giustizia alternative all'egualitarismo semplice insieme alle motivazioni morale sottostanti. In questa scheda, consideriamo la loro definizione matematica. Come vedremo nelle righe successive, per le prime quattro teorie, la giustizia è funzione del modo in cui l'utilità è distribuita tra i gruppi. Tali teorie si differenziano dall'utilitarismo, che considera la somma delle utilità trascurando del tutto come essa è distribuita tra i diversi gruppi.

La domanda a cui vogliamo rispondere ora è: come dovrebbe essere distribuita l'utilità tra gruppi (definiti dall'attributo sensibile)?

Egualitarismo

Nel caso dell'egualitarismo l'equità si ha se gli individui di entrambi i gruppi ricevono in media la stessa utilità dalla regola decisionale. L'uguaglianza tra gruppi è apprezzata in quanto tale. Dati n gruppi distinti, che chiameremo G_1, G_2, ..., G_n, possiamo riassumere il principio grazie alle seguenti uguaglianze:

$$\mathbb{E}_{SD}(G_1) = \mathbb{E}_{SD}(G_2) = \cdots = \mathbb{E}_{SD}(G_n).$$

Punteggio di equità. Come si misura il punteggio di equità? In questo caso ciò equivale a chiedersi: quanto sono vicine le utilità medie di due gruppi ad essere uguali?

Consideriamo, per semplicità, solo due gruppi G_1 e G_2, allora il punteggio di equità nel caso dell'egualitarismo S_{Eg} è dato da

$$S_{Eg} = |\mathbb{E}_{SD}(G_1) - \mathbb{E}_{SD}(G_2)|.$$

Maximin

Nel caso del principio di maxmin l'equità si ha se si utilizza la regola migliore (nel senso che produce la maggiore utilità) r' per il gruppo più svantaggiato.

Nella Scheda 3, abbiamo visto che la scelta di una determinata regola decisionale (in questo caso una soglia di probabilità) determina la distribuzione dei numeri nella tabella di decisione, che può essere diversa per ciascun gruppo G. Nella Scheda 8, abbiamo mostrato che l'utilità media di ciascun gruppo può essere espressa in funzione dei numeri della tabella di utilità del Soggetto alla Decisione. Possiamo quindi sintetizzare questa dipendenza sia dal gruppo che dalla regola con una doppia notazione funzionale:

$$\mathbb{E}_{SD}(G)(r).$$

Grazie a questa notazione, il criterio maxmin si può esprimere tramite la seguente formula:

$$\mathbb{E}_{SD}(G_{\text{svantaggiato}})(r') = \max_{\text{(tutte le regole } r)} \mathbb{E}_{SD}(G_{\text{svantaggiato}})(r),$$

Punteggio di equità Come si misura il punteggio di equità? In questo caso ciò equivale a chiedersi: qual è l'utilità media più bassa del gruppo[5]?

$$S_{Mm} = \mathbb{E}_{SD}(G_{\text{svantaggiato}})(r).$$

Prioritarismo

Nel caso del prioritarismo l'equità si ottiene se l'utilità aggregata dei gruppi è massimizzata dalla regola decisionale, con l'utilità del gruppo più sfavorito pesata maggiormente rispetto alle utilità degli altri gruppi.

Introduciamo un parametro $k > 1$ che rappresenta il peso sopracitato da assegnare all'utilità del gruppo più svantaggiato[6]. Nel caso di due gruppi, $G_{\text{svantaggiato}}$, $G_{\text{avvantaggiato}}$ vogliamo massimizzare la seguente quantità:

$$\hat{\mathbb{E}}_{SD} = k \cdot \mathbb{E}_{SD}(G_{\text{svantaggiato}}) + \mathbb{E}_{SD}(G_{\text{avvantaggiato}}),$$

cioè trovare la regola r' tale che

$$\hat{\mathbb{E}}_{SD}(r') = \max_{\text{tutte le regole } r} \hat{\mathbb{E}}_{SD}(r).$$

[5] Più l'utilità è alta, maggiore è il punteggio di equità.

[6] Poiché $k > 1$ l'utilità del gruppo più svantaggiato conta di più di quella degli altri gruppi.

Punteggio di equità. Come si misura il punteggio di equità? In questo caso ciò equivale a chiedersi: qual è l'utilità aggregata con il gruppo peggiore che ha un peso maggiore nella aggregazione?

$$S_{Pr} = \hat{\mathbb{E}}_{SD}(r).$$

Sufficientismo

Nel caso del sufficientismo l'equità si raggiunge se tutti i gruppi hanno un'utilità media superiore alla soglia definita. Le disuguaglianze sono accettabili se ogni gruppo è al di sopra della soglia definita. In formule, fissata una soglia (cioè un numero) t, abbiamo

$$\mathbb{E}_{SD}(G) \geq t.$$

Punteggio di equità. Come si misura il punteggio di equità? In questo caso ciò equivale a chiedersi: quanti gruppi sono al di sopra della soglia?

$$S_{Su} = \#\{G \text{ tali che } \mathbb{E}_{SD}(G) \geq t\},$$

cioè il numero di gruppi G la cui utilità media è sopra la soglia fissata. Ricordiamo che nella Scheda 7 abbiamo considerato l'utilitarismo alla stregua di una teoria della giustizia. Questa teoria, tuttavia, non considera rilevante il modo in cui l'utilità è distribuita tra i diversi gruppi.

Utilitarismo

Nella teoria dell'utilitarismo la scelta più equa è quella che massimizza il benessere medio della popolazione, senza distinguere a quale gruppo appartiene. In particolare abbiamo:

$$\mathbb{E}_{SD}(r') = \max_{\text{tutte le regole } r} \mathbb{E}_{SD}(r),$$

dove non è presente alcuna dipendenza dal gruppo in quanto il valore di utilità viene calcolato per la popolazione nel suo complesso (come se fosse un unico gruppo).

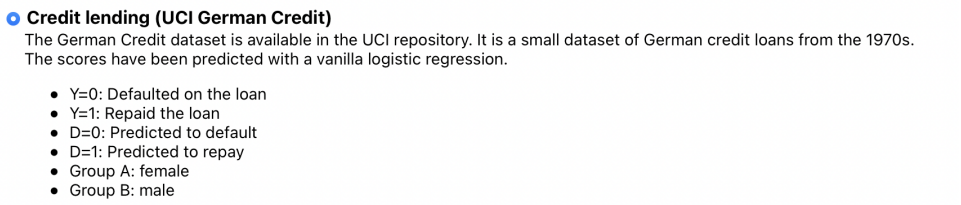
Punteggio di equità. Come si misura il punteggio di equità? In questo caso ciò equivale a chiedersi: quanto è alta l'utilità media?

$$S_{Ut} = \mathbb{E}_{SD}(r).$$

10. Laboratorio: un metodo di discussione della diseguaglianza tra gruppi

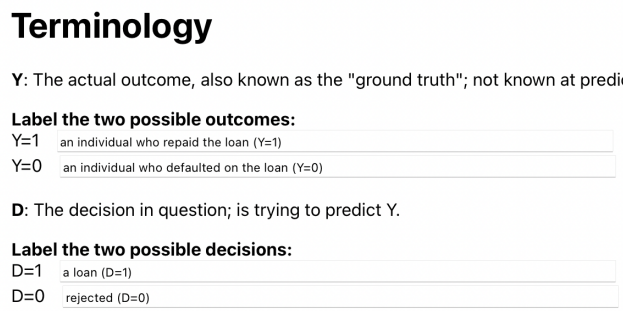
Il gruppo di ricerca Socially acceptable and fair AI con sede a Zurigo ha creato un software (in inglese) denominato FairnessLab [7] che facilita la analisi della giustizia algoritmica tra gruppi, in un caso concreto. Caratteristica del lab è che richiede allo user di specificare alcuni parametri, i quali corrispondono ad assunti di valore, come vedremo. Consideriamo l'esempio della banca che usa un algoritmo per decidere se concedere un mutuo a un cliente.

Il primo passo consiste nel selezionare un set di dati che contiene le variabili del nostro modello, cioè Y, D, e distinguere il gruppo a cui appartengono i SD (Soggetti alla Decisione). In questo caso indicheremo il gruppo con la variabile A.



Quando testiamo l'algoritmo attraverso i dati dei precedenti clienti possiamo distinguere due sottogruppi di ciascun gruppo di SD in base a quello che è stato l'esito effettivo del prestito per ciascun individuo, cioè il valore della variabile Y.

[7] Disponibile gratuitamente sul sito: https://joebaumann.github.io/FairnessLab/#/.

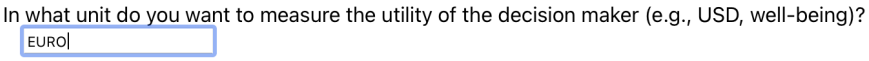


Il nostro dataset inoltre permette di calcolare la probabilità che ciascun individuo nel dataset ripagherà il prestito. Come mostrato dall' esercizio esplicativo questa probabilità permette di determinare la ripartizione della popolazione nei quattro riquadri della tabella di decisione (Scheda 3).

I prossimi elementi, invece, devono essere definiti da coloro che intraprendono la valutazione di equità.

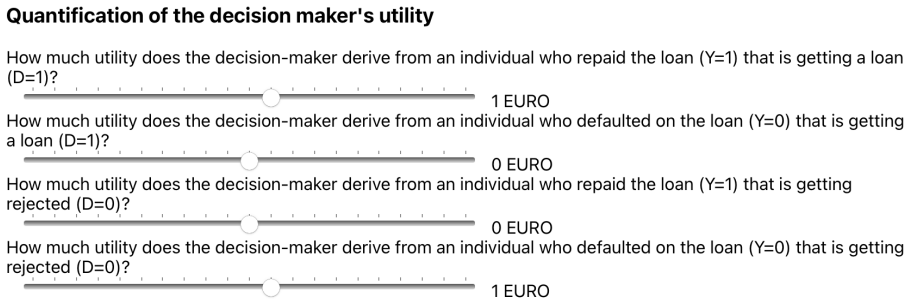
Inizialmente, dobbiamo stabilire il costo dell'equità dal punto di vista della banca. Per fare ciò, (attraverso il modello matematico esposto nella Scheda 3) occorre fornire al modello un parametro che non è pre-impostato: il beneficio derivante alla banca per ciascun prestito concesso e ciascun prestito negato, sia nel caso in cui il prestito venga restituito sia nel caso in cui ciò non accada. Nella realtà, il beneficio in questione dipende dall'entità del prestito e da diverse variabili di mercato e persino macro-economiche, come le aspettative relative ad aumenti del tasso di interesse da parte della banca centrale. A scopo illustrativo, noi immaginiamo di avere a che fare con prestiti tutti della stessa entità e che il tasso di interesse richiesto al cliente sia necessariamente lo stesso per tutti.

La prima cosa che il programma richiede è di definire un'unità di misura in cui valutare il beneficio ottenuto dalla banca.



In questo caso, supponiamo che il beneficio della banca possa essere valutato in Euro.

Occorre ora fornire al programma i valori specifici che corrispondono a quelli della tabella di utilità della Scheda 2. Questi valori corrispondono al beneficio o danno effettivo di ciascuna decisione, nel momento in cui si realizza l'esito Y.



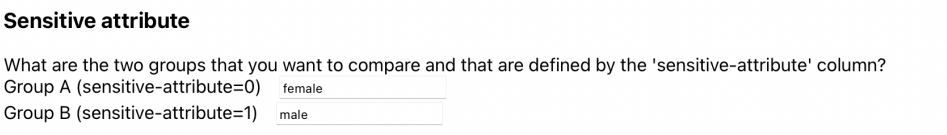
A questo punto abbiamo completato la parte della procedura che serve a valutare il costo dell'equità dal punto di vista dello scopo primario del decisore. Passiamo ora alla quantificazione della giustizia del nostro algoritmo, quella che nella scheda 8 abbiamo chiamato "misura della giustizia", e che in questo software è chiamato "fairness score", cioè punteggio di equità.

Per fare ciò, è necessario passare in rassegna tutti i parametri normativi del nostro modello, che sono:

- i gruppi che consideriamo socialmente salienti (Scheda 4);
- il differenziatore delle pretese (che modella le diverse teorie della giustizia della Scheda 7);
- il valore di utilità per i soggetti alla decisione (Scheda 8);
- lo schema di giustizia (che determina il punteggio di equità come definito nella Scheda 9).

I gruppi

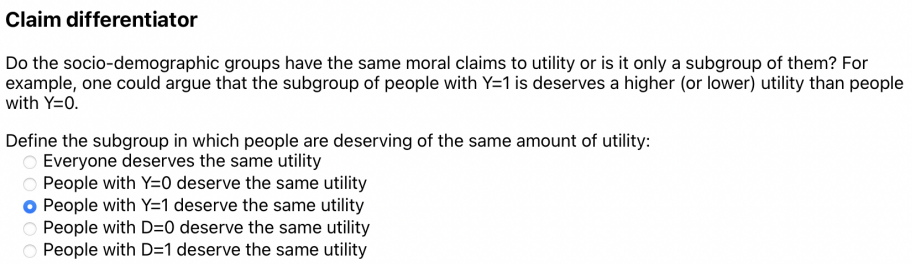
Il primo elemento consiste nella definizione dei gruppi socialmente salienti (Scheda 4) che intendiamo comparare. Per questo dataset la nostra scelta è limitata, cioè possiamo distinguere solo i due sessi, maschile e femminile:



Il differenziatore delle pretese morali

Il secondo elemento corrisponde alla questione se alcune diseguaglianze sono giustificate dalle diverse qualità (moralmente rilevanti) degli individui coinvolti nelle

decisioni, posta dalla Scheda 5. Tale scheda distingue tre tipi di teorie della diseguaglianza, la teoria basata sui bisogni, quella basata sul merito (inteso come contributo alla cooperazione), e quella basata sulla responsabilità. Nel modello da noi implementato nel software utilizziamo un parametro il cui valore è determinato dallo user, il *claim differentiator*, che permette allo user di stabilire se intende utilizzare una teoria basata sui bisogni, i contributi, o le responsabilità, a seconda di quanto il contesto richiede. Possiamo tradurre l'espressione "claim differentiator" in italiano come "differenziatore delle pretese". Il differenziatore delle pretese è quell'elemento (che può corrispondere, a seconda dei casi, a un bisogno, un contributo, o una responsabilità) che giustifica un diverso trattamento per i diversi individui.



Nell'esempio della banca, nessuno dei dati a nostra disposizione consente di distinguere gli individui in base ai loro diversi bisogni o al loro diverso grado di responsabilità. Tuttavia abbiamo un dato, il valore della variabile da predire Y (di cui conosciamo il valore nel dataset storico) che può essere interpretato come moralmente rilevante, in base a una teoria del merito inteso come contributo.

Possiamo quindi considerare due ipotesi opposte sulla natura della giustizia in questo contesto:

(a) tutti i clienti meritano lo stesso beneficio, a prescindere dal fatto che in futuro ripagheranno il loro debito, oppure no;

(b) sono giustificate le diseguaglianze che riflettono il diverso contributo degli individui allo schema di cooperazione, cioè i clienti che in futuro restituiranno il debito dovrebbero idealmente ricevere un trattamento diverso dagli altri.

I casi implementati in questo esempio permettono di selezionare l'ipotesi (b). Questa ipotesi corrisponde alla teoria della giustizia secondo cui tutti i clienti che ripagheranno il debito ($Y = 1$) meritano lo stesso beneficio, perché sono tutti uguali dal punto di vista del loro contributo alla stabilità finanziaria dell'istituto che concede il prestito. Per coloro che, invece, non ripagheranno il debito ($Y = 0$), non è doveroso garantire le stesse aspettative. Dunque la caratteristica Y, ripagare

il debito, determina due classi diverse dal punto di vista di una teoria morale del merito, inteso come contributo.

Selezionando $Y = 1$ generiamo un punteggio di equità che compara l'utilità media del gruppo "uomini" inteso come "uomini che ripagano il debito" (cioè l'intersezione del gruppo "uomini" con il gruppo $Y = 1$) e del gruppo "donne" inteso come "donne che ripagano il debito" (cioè l'intersezione del gruppo "donne" con il gruppo $Y = 1$). L'utilità attesa generata dall'algoritmo per coloro, uomini o donne, che *non* ripagano il debito con la banca viene ignorata se selezionamo $Y = 1$ come "claim differentiator".

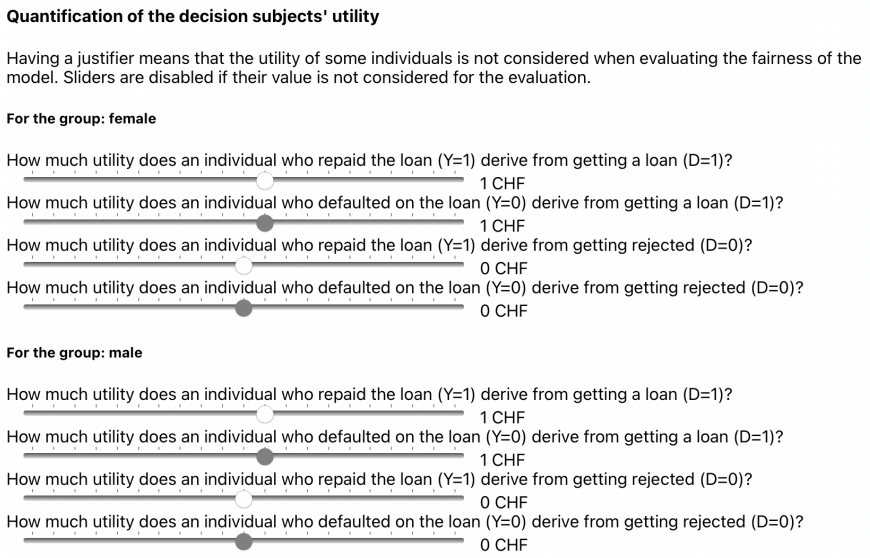
Il valore di utilità per il soggetto alla decisione

Il passo seguente serve a fornire al software tutti i parametri necessari per calcolare il benessere atteso dei diversi gruppi, calcolato in base al modello presentato nella Scheda 9.

Ciò che il programma richiede al valutatore è di specificare i valori della tabella di utilità del soggetto alla decisione (Scheda 9). È difficile valutare questi valori in modo empirico. La nostra ipotesi è che tra persone ragionevoli sia possibile accordarsi su una teoria del valore delle decisioni per l'individuo medio di un determinato gruppo e che questo genere di "teorie del senso comune" possano fornire una base sensata per una valutazione approssimativa della giustizia algoritmica, a dispetto della loro inesorabile soggettività.

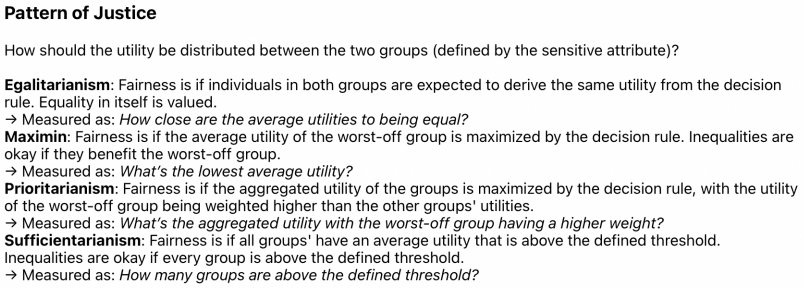
Come spiegato nella Scheda 9, il nostro modello matematico permette di modellare la possibilità in cui il beneficio che il SD riceve da ciascun tipo di decisione ($D = 0$ o $D = 1$) sia diverso a seconda dell'esito ($Y = 0$ o $Y = 1$) effettivo della decisione. In questo caso, ciò equivale a modellare la possibilità in cui il beneficio per il cliente che riceve un prestito dipenda dalla capacità reale di quel cliente di ripagarlo.

Si noti che la nostra cornice analitica consente inoltre di assegnare un diverso vantaggio in corrispondenza delle stesse decisioni D, e esiti Y, diversi per i diversi gruppi (in questo caso, uomini e donne). In altre parole, è possibile modellare la possibilità che il beneficio di ricevere un prestito sia diverso a seconda del sesso del cliente.



Schema della giustizia

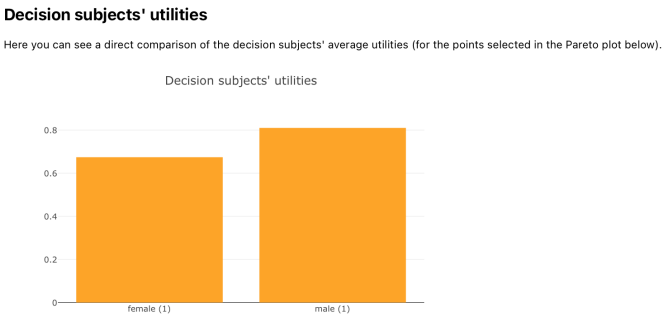
Infine, la nostra cornice analitica richiede che venga identificata la concezione della giustizia che si considera come valida dal punto di vista normativo, scegliendo tra quattro concezioni della giustizia diverse:



Una volta forniti questi input, il programma calcola l'utilità attesa sia per il decisore (la banca, in questo saggio) sia per i gruppi rilevanti (gli uomini che ripagano il loro debito e le donne che ripagano il loro debito, in questo caso). La comparazione delle utilità attese per i gruppi rilevanti determina un punteggio di equità, che sarà diverso a seconda dello schema della giustizia che abbiamo selezionato come mostrato dalla Scheda 9.

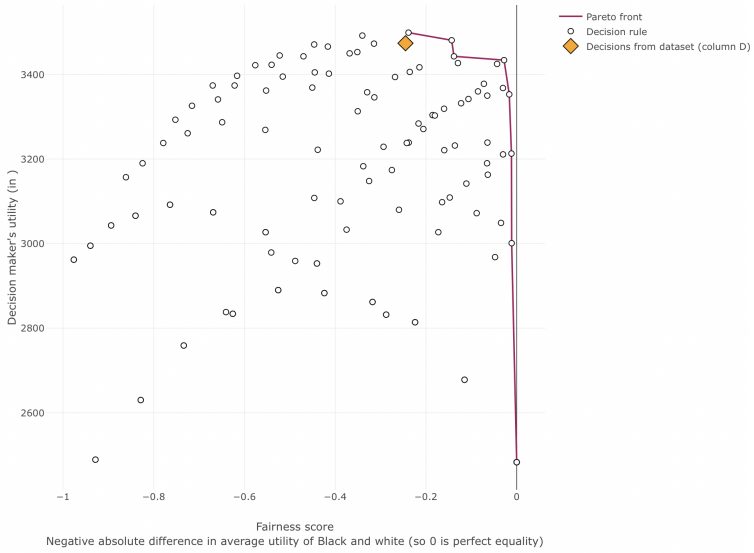
Visualizzazioni del fairness lab

Il programma poi consente di rappresentare le utilità medie di ciascun gruppo (ad esempio, il gruppo degli uomini e quello delle donne che sono effettivamente capaci a restituire il loro debito).



Qui sotto, invece, il "punteggio di equità". Questo valore è tanto più alto, quanto maggiore è l'allineamento degli effetti della decisione con la concezione di giustizia da noi impostata come valida dal punto di vista normativo. Ad esempio, se la concezione della giustizia selezionata è l'egualitarismo, questo punteggio è massimo quando l'utilità media dei due gruppi è esattamente la stessa.

Infine, il programma rappresenta su un piano cartesiano l'utilità attesa per il decisore (asse delle ordinate) e il punteggio di equità (asse delle ascisse) che corrisponde a ogni possibile regola di decisione considerata. In questo caso, le regole di decisione considerate sono regole soglia, ad esempio "concedi il prestito se la probabilità di restituirlo è più alta di 0.7". Ogni punto nel grafico qui sotto corrisponde a una regola soglia diversa. Il grafico più sotto si ottiene considerando 11 soglie per ciascun gruppo. Poiché consideriamo le regole di decisione che usano soglie diverse per i diversi gruppi (in questo caso, per uomini e donne) i punti rappresentati sono molti di più di 11, in quanto rappresentano tutte le combinazioni possibili di 11 soglie diverse per gli uomini e 11 soglie diverse per le donne.



Il concetto e la rappresentazione della Pareto-ottimalità

Rappresentando in questo modo l'utilità del decisore e il punteggio di equità, possiamo distinguere le regole di decisione (tra quelle considerate nel grafico) che sono Pareto ottimali da quelle che non lo sono. Le combinazioni Pareto ottimali sono quelle che non consentono di migliorare l'equità delle nostre scelte senza ridurre il benessere atteso del decisore e, viceversa, che non consentono di migliorare il benessere atteso del decisore senza ridurre il punteggio di equità. Dal punto di vista grafico, le combinazioni Pareto ottimali sono quelle che risultano nel bordo esterno verso l'alto e verso destra del grafo Cartesiano.

Si noti che quando esaminiamo una regola di decisione che non è Pareto ottimale, possiamo facilmente scartarla a favore di una regola di decisione che sia superiore a questa da qualche punto di vista (ad esempio, più equa, oppure più conveniente per il decisore) senza che sia peggiore da alcun punto di vista. La scelta di un punto che non stia sulla frontiera di Pareto risulta quindi, coerentemente con le ipotesi iniziali, irrazionale.

Quando esaminiamo l'una con l'altra diverse regole di decisione che siano Pareto ottimali risulta invece molto difficile giustificare una scelta semplicemente sulla base del modello di analisi e delle teorie etiche considerate fino a questo punto. Si noti che la nostra cornice di analisi non fornisce alcun "meta-valore" che permetta di stabilire se preferire una regola più conveniente per il decisore o una che sia più equa. Convenienza per il decisore ed equità sono modellati come due obiettivi che non possono essere ricondotti ad alcuna matrice comune. Ci viene quindi richiesto di esprimere una preferenza per un po' di equità in più oppure un po' di convenienza

in più. Questa preferenza è modellata come se fosse una preferenza bruta, non soggetta ad ulteriore analisi o giustificazione morale, cioè come un gusto soggettivo. Questo però non è il modo più corretto di pensare al ruolo della moralità nelle decisioni algoritmiche. Piuttosto, il limite della formalizzazione da noi adottato riflette considerazioni pragmatiche e di semplicità formale. Sperabilmente, gli stakeholder potranno giustificare le preferenze per punti Pareto ottimali *specifici* con delle buone buone ragioni di natura morale *e prudenziale*. Ad esempio, la banca potrebbe sollevare l'obiezione, rispettto alla distribuzione che raggiunge il più alto punteggio di equità, che essa non è compatibile con gli obblighi di sostenibilità finanziaria verso il regolatore. Quindi questo potrebbe giustificare la scelta di un punto (Pareto ottimale) che sacrifichi un po' di equità trai gruppi migliorando la performance finanziaria del decisore. Tuttavia, il nostro modello non fornisce alcuna guida esplicita per questo tipo di scelta, se non quello di evitare le regole che non sono Pareto ottimali. Lasciamo all'utilizzatore del lab l'onere di sviluppare il discorso etico per valutare gli spostamenti lungo la frontiera di Pareto, in assenza di una guida esplicita da parte del nostro approccio.

Ringraziamenti

Si ringraziano Nicolò Cangiotti per la consulenza editoriale e l'editing e Domenico Brunetto per i preziosi suggerimenti che hanno indubbiamente migliorato il lavoro. Queste schede si basano sulla teoria sviluppata da Michele Loi, Christoph Heitz, Joachim Baumann, e Corinna Hertweck e descritta in due working papers [4, 12] e in un libro dal titolo "Just Machines? A philosophical and practical approach to fairness in machine learning", di prossima uscita. Le schede 3 e 4 si basano sulle dispense di Christoph Heitz per il corso su "Social impact of AI and algorithmic fairness" della Università di scienze applicate di Zurigo, rielaborate da Nicolò Cangiotti.

Bibliografia

[1] Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner. Machine bias. *ProPublica*, 23(2016):139–159, 2016.

[2] Richard J. Arneson. Equality and Equality of Opportunity for Welfare. *Philosophical Studies*, 56(1):77–93, 1989.

[3] Richard J. Arneson. Luck Egalitarianism and Prioritarianism. *Ethics*, 110(2):339–349, 2000.

[4] Joachim Baumann, Corinna Hertweck, Michele Loi, and Christoph Heitz. Distributive Justice as the Foundational Premise of Fair ML: Unification, Extension, and Interpretation of Group Fairness Metrics, 2022. arXiv:2206.02897.

[5] Tim Brennan, William Dieterich, and Beate Ehret. Evaluating the predictive validity of the compas risk and needs assessment system. *Criminal Justice and Behavior*, 36(1):21–40, 2009.

[6] Huub Brouwer and Thomas Mulligan. Why not be a desertist? *Philosophical Studies*, June 2018.

[7] Sam Corbett-Davies, Emma Pierson, and Sharad Goel. A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. *Washington Post*, 2021.

[8] Ronald Dworkin. What is Equality? Part 2: Equality of Resources. *Philosophy and Public Affairs*, 10(4):283–345, 1981.

[9] Henry Frankfurt. Equality as a Moral Ideal. *Ethics*, 98:21–43, 1987.

[10] Elena Granaglia. *Uguaglianza di opportunità: Sì, ma quale?* Editori Laterza, 2022.

[11] Hoda Heidari, Michele Loi, Krishna P. Gummadi, and Andreas Krause. A Moral Framework for Understanding Fair ML Through Economic Models of Equality of Opportunity. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pages 181–190, New York, NY, USA, 2019. ACM. event-place: Atlanta, GA, USA.

[12] Corinna Hertweck, Joachim Baumann, Michele Loi, Eleonora Viganò, and Christoph Heitz. A Justice-Based Framework for the Analysis of Algorithmic Fairness-Utility Trade-Offs, 2022. arXiv:2206.02891.

[13] Corinna Hertweck, Christoph Heitz, and Michele Loi. On the Moral Justification of Statistical Parity. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pages 747–757, New York, NY, USA, March 2021. Association for Computing Machinery.

[14] David Ingold and Spencer Soper. Amazon Doesn't Consider the Race of Its Customers. Should It? in: https://www.bloomberg.com/graphics/2016-amazon-same-day/, 2016.

[15] David Miller. *Principles of Social Justice*. Harvard University Press, Cambridge, Mass, 1999.

[16] Derek Parfit. Equality and Priority. In *Debates in Contemporary Political Philosophy: An Anthology*, pages 115–132. Routledge, London, 2003.

[17] John Rawls. *A Theory of Justice*. Harvard University Press, Cambridge, MA, 1971.

[18] Nicola Riva. *Eguaglianza delle opportunità*. Aracne, 2011.

[19] Liam Shields. *Just enough: sufficiency as a demand of justice*. Edinburgh University Press, Edinburgh, 2016.

[20] David Wiggins. Claims of Need. In *Needs, Values, Truth.*, pages 1–58. Oxford University Press, Oxford, 1987.