This week we're building up our skills at inventing useful probability models.

MONDAY

- linear models & feature design

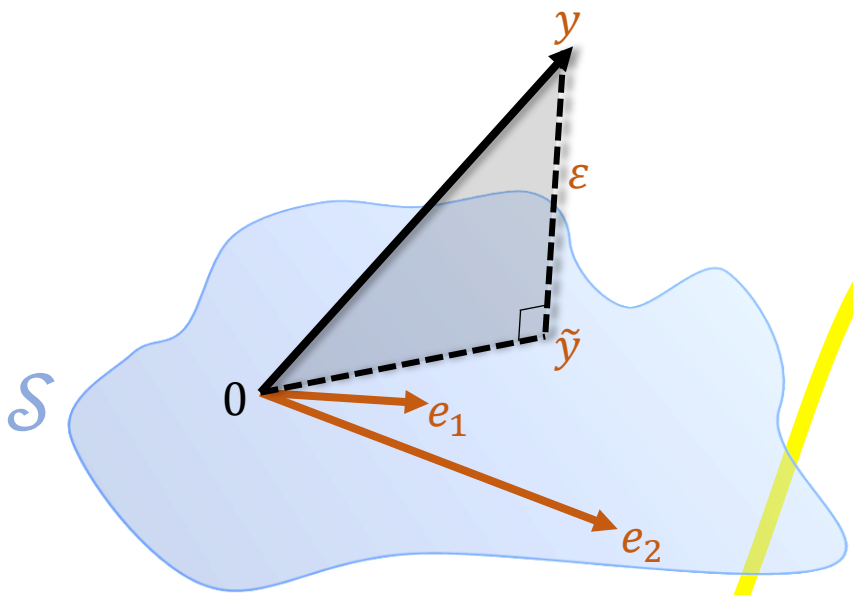
(to quickly turn our ideas into easy-to-fit models)

WEDNESDAY

- "debugging" models
- linear algebra interpretation

FRIDAY

- identifiability of parameters
- the link between least squares and likelihood



What does "closest" even mean?
It means: find $\tilde{y} \in \mathcal{S}$ to minimize $\|y - \tilde{y}\|$

i.e. to minimize $\sqrt{\sum_i \varepsilon_i^2}$ where $\varepsilon = y - \tilde{y}$.

The minimization is over all $\tilde{y} \in \mathcal{S}$,
i.e. over all linear combinations of $\{e_1, \ldots, e_K\}$.

GEOMETRY OF PROJECTIONS

Let $\mathcal{S}$ be the span of $\{e_1, \ldots, e_K\}$.

What's the closest we can get to y, while staying in $\mathcal{S}$?

What does "best approximation" mean?
It means: find $\beta_1, \ldots, \beta_K$ to minimize the mean square error,

$$\frac{1}{\#\text{datapoints}} \sum_i \varepsilon_i^2$$

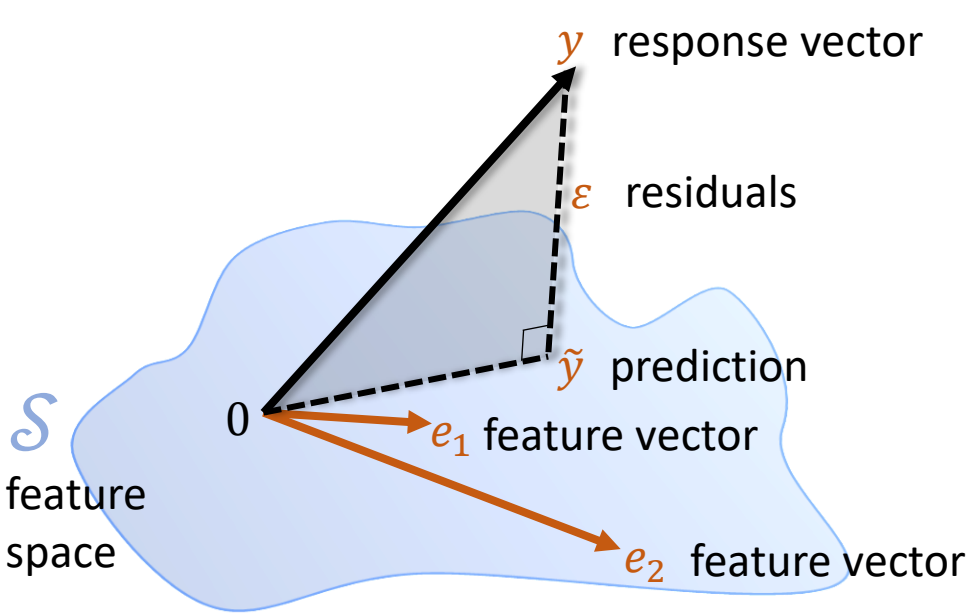
where $\varepsilon = y - (\beta_1 e_1 + \cdots + \beta_K e_K)$

LINEAR MODELLING / LEAST SQUARES ESTIMATION

Given features $\{e_1, \ldots, e_K\}$ let's approximate $y \approx \beta_1 e_1 + \cdots + \beta_K e_K$.

What parameters give us the best approximation?

These are exactly the same thing!



The span of $\{e_1, \ldots, e_K\}$ is called the *feature space*

- Fitting a linear model ≡ projecting the data y onto the feature space $\mathcal{S}$
- The fitted parameters β_k are the "coordinates" of the predicted response, $\tilde{y} = \sum_k \beta_k e_k$
- If two models have different features but the same feature space, they'll make exactly the same prediction $\tilde{y}$ (but of course have different β_k)
- If the feature vectors are linearly independent, then the β_k are uniquely identified. Otherwise, the parameters have no intrinsic meaning, and we say they're *confounded*.

Exercise 2.6.2 (Contrasts)

For the dataset below, of measurements from two groups A and B, discuss the differences between these three models:

$$y \approx \alpha 1_{g=A} + \beta 1_{g=B} \quad \text{(M1)}$$
$$y \approx \alpha' + \beta' 1_{g=B} \quad \text{(M2)}$$
$$y \approx \alpha'' + \beta'' 1_{g=A} + \gamma'' 1_{g=B} \quad \text{(M3)}$$

g	y
A	0.5
A	1.9
B	3.5
B	1.1
B	2.3

M1: features are $\begin{bmatrix}1\\1\\0\\0\\0\end{bmatrix}$ and $\begin{bmatrix}0\\0\\1\\1\\1\end{bmatrix}$

M2: features are $\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}$ and $\begin{bmatrix}0\\0\\1\\1\\1\end{bmatrix}$

These span the same space, so M1 and M2 make identical predictions on the dataset (though they use different parameterizations).

M3: features are $\begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix}$ and $\begin{bmatrix}1\\1\\0\\0\\0\end{bmatrix}$ and $\begin{bmatrix}0\\0\\1\\1\\1\end{bmatrix}$

These are linearly dependent, so the parameters are not identifiable — there are multiple equivalent ways of writing out the same response vector,

$$\vec{y} \approx 1.2\, 1_{\vec{y}=A} + 2.3\, 1_{\vec{y}=B}$$
$$\approx \vec{1} + 0.2\, 1_{\vec{y}=A} + 1.3\, 1_{\vec{y}=B}$$
$$\approx 2.3\, \vec{1} - 1.1\, 1_{\vec{y}=A}.$$

§2.6 Interpreting parameters

- Check if the features are linearly dependent.
  If they are, the parameters have no intrinsic meaning.
  We say the features are *confounded*, and the parameters are *non-identifiable.*
- Write out the predicted response for a few typical / representative datapoints.
  This helps see what the parameters mean.
- It may be useful to use different features (with the same feature space).
  This doesn't change the model's predictions, but it does change the parameterization. It lets us extract the *contrasts* i.e. *comparisons* we're interested in.

Sign in

The Guardian

News website of the year

Stop and search

This article is more than **3 years old**

Met police 'disproportionately' use stop and search powers on black people

London's minority black population targeted more than white population in 2018 – official figures

Do I trust this finding?

- Proportionate to what?
- Is it cherry picking?
- Why the scare quotes?

Sign in

The Guardian

News website of the year

Stop and search

This article is more than **3 years old**

Met police 'disproportionately' use stop and search powers on black people

London's minority black population targeted more than white population in 2018 – official figures

Can I set up a model with a parameter that *measures* the quantity I'm interested in?

force	Date	LatLng	Object of search	Gender	Age range	Officer-defined ethnicity	Outcome
cambridgeshire	2023-08-31 15:44:04+00:00	(52.43,-0.142)	Controlled drugs				A no further action disposal
cambridgeshire	2023-08-31 15:35:41+00:00	(52.43,-0.142)	Firearms	Male	25-34	White	Khat or Cannabis warning
cambridgeshire	2023-08-31 14:44:04+00:00	(52.43,-0.142)	Firearms	Male	25-34	White	Khat or Cannabis warning
cambridgeshire	2023-08-31 03:44:14+00:00	(52.58,-0.244)	Offensive weapons	Male		Other	A no further action disposal
cambridgeshire	2023-08-31 02:34:16+00:00	(52.59,-0.247)	Controlled drugs	Male	25-34	White	Arrest
cambridgeshire	2023-08-31 02:27:10+00:00	(52.21,0.124)	Controlled drugs	Male	18-24	White	A no further action disposal
cambridgeshire	2023-08-30 22:28:13+00:00	(52.45,-0.117)	Controlled drugs	Female	over 34	White	A no further action disposal
cambridgeshire	2023-08-30 20:24:13+00:00	(52.32,-0.0708)	Controlled drugs	Male	10-17	White	Summons / charged by post
cambridgeshire	2023-08-30 14:26:58+00:00	(52.57,-0.24)	Controlled drugs	Male	over 34	Asian	A no further action disposal
cambridgeshire	2023-08-30 14:13:45+00:00	(52.57,-0.24)	Controlled drugs	Male	25-34	Black	Arrest

Log of England+Wales stop-and-search incidents, from the UK home office https://data.police.uk/

Example 2.6.4

The UK Home Office makes available a dataset of police stop-and-search incidents. We wish to investigate whether there is racial bias in police decisions to stop-and-search. Consider the linear model

$$y_i \approx \alpha + \beta_{\texttt{eth}_i}$$

where $\texttt{eth}_i$ is the officer-defined ethnicity for record i, and y_i records the outcome: $y_i = 1$ if the police found something, 0 otherwise.

a) Write this as a linear model using one-hot coding.

b) Are the parameters identifiable? If not, rewrite the model so that they are.

c) Does the model suggest there is racial bias in policing actions?

Example 2.6.4

The UK Home Office makes available a dataset of police stop-and-search incidents. We wish to investigate whether there is racial bias in police decisions to stop-and-search. Consider the linear model

$$y_i \approx \alpha + \beta_{\texttt{eth}_i}$$

where $\texttt{eth}_i$ is the officer-defined ethnicity for record i, and y_i records the outcome: $y_i = 1$ if the police found something, 0 otherwise.

For records with $\texttt{eth}_i = k$, the average response is

$$\frac{\sum_{i:\text{eth}_i=k} y_i}{|\{i: \text{eth}_i = k\}|}$$

This is just #finds/#stops, i.e.

$$\mathbb{P}(\text{find something})$$

For records with $\texttt{eth}_i = k$, the predicted response is $\alpha + \beta_k$.

Actual responses y_i may be above or below this prediction, but their average is $\alpha + \beta_k$.

This model is thus proposing: $\mathbb{P}(\text{find something}) \approx \alpha + \beta_k$ for a person in ethnic group k

If $\beta_k < 0$, that means $\mathbb{P}(\text{find something})$ is low compared to other ethnic groups, i.e. the police are stopping relatively more innocent people.

Example 2.6.4

The UK Home Office makes available a dataset of police stop-and-search incidents. We wish to investigate whether there is racial bias in police decisions to stop-and-search. Consider the linear model

$$y_i \approx \alpha + \beta_{\text{eth}_i} \quad \text{i.e.} \quad \vec{y} \approx \alpha \vec{1} + \sum_k \beta_k 1_{\text{eth}=k}$$

e.g. for eth = mixed, $y \approx \alpha + \beta_{\text{Mi}}$

where eth_i is the officer-defined ethnicity for record i, and y_i records the outcome: $y_i = 1$ if the police found something, 0 otherwise.

a) Write this as a linear model using one-hot coding.

$$y \approx \alpha 1 + \underset{\text{Asian}}{\beta_{\text{As}} 1_{\text{eth}=\text{As}}} + \underset{\text{Black}}{\beta_{\text{Bl}} 1_{\text{eth}=\text{Bl}}} + \underset{\text{Mixed}}{\beta_{\text{Mi}} 1_{\text{eth}=\text{Mi}}} + \underset{\text{Other}}{\beta_{\text{Oth}} 1_{\text{eth}=\text{Oth}}} + \underset{\text{White}}{\beta_{\text{Wh}} 1_{\text{eth}=\text{Wh}}}$$

```
1   ethnicity_levels = np.unique(eth)
2   eth_onehot = [np.where(eth==k,1,0) for k in ethnicity_levels]
3
4
5   model = sklearn.linear_model.LinearRegression()
6   model.fit(np.column_stack(eth_onehot), y)
7   α,βs = model.intercept_, model.coef_
8
9   print(f'α = {α}')
10  for k,β in zip(ethnicity_levels, βs):
11      print(f'β[{k}] = {β}')
```

```
α = -34037792910.00365
β[Asian] = 34037792910.26522
β[Black] = 34037792910.265717
β[Mixed] = 34037792910.2939
β[Other] = 34037792910.260
β[White] = 34037792910.26
```

Example 2.6.4

The UK Home Office makes available a dataset of police stop-and-search incidents. We wish to investigate whether there is racial bias in police decisions to stop-and-search. Consider the linear model

$$y_i \approx \alpha + \beta_{\mathrm{eth}_i}$$

where eth_i is the officer-defined ethnicity for record i, and y_i records the outcome: $y_i = 1$ if the police found something, 0 otherwise.

a) Write this as a linear model using one-hot coding.

b) Are the parameters identifiable? If not, rewrite the model so that they are.

$\vec{y} \approx \alpha \vec{1} + \sum_k \beta_k 1_{\vec{\mathrm{eth}}=k}$

(a) $$y \approx \alpha 1 + \beta_{\mathrm{As}} 1_{\mathrm{eth}=\mathrm{As}} + \beta_{\mathrm{Bl}} 1_{\mathrm{eth}=\mathrm{Bl}} + \beta_{\mathrm{Mi}} 1_{\mathrm{eth}=\mathrm{Mi}} + \beta_{\mathrm{Oth}} 1_{\mathrm{eth}=\mathrm{Oth}} + \beta_{\mathrm{Wh}} 1_{\mathrm{eth}=\mathrm{Wh}}$$

(b) Not identifiable, since the features are linearly dependent: $\vec{1} = 1_{\vec{\mathrm{eth}}=\mathrm{As}} + 1_{\vec{\mathrm{eth}}=\mathrm{Bl}} + \cdots + 1_{\vec{\mathrm{eth}}=\mathrm{Wh}}$.

To get linear independence:

— we could ditch the $\vec{1}$ feature, $\vec{y} \approx \beta_{\mathrm{As}} 1_{\vec{\mathrm{eth}}=\mathrm{As}} + \cdots + \beta_{\mathrm{Wh}} 1_{\vec{\mathrm{eth}}=\mathrm{Wh}}$

— or we could ditch one of the others, $\vec{y} \approx \alpha + \beta_{\mathrm{Bl}} 1_{\vec{\mathrm{eth}}=\mathrm{Bl}} + \cdots + \beta_{\mathrm{Wh}} 1_{\vec{\mathrm{eth}}=\mathrm{Wh}}$

Example 2.6.4

c) Does this model suggest there is racial bias in policing actions?

First, let's interpret the parameters of the rewritten model. $\hat{y} = \alpha + \beta_{Bl} 1_{eth=Bl} + \cdots + \beta_{Wh} 1_{eth=Wh}$

For a person with eth = As predicted $y = \alpha$

eth = Bl	$\alpha + \beta_{Bl}$
eth = Mi	$\alpha + \beta_{Mi}$
eth = Oth	$\alpha + \beta_{Oth}$
eth = Wh	$\alpha + \beta_{Wh}$

These β_k' measure <u>differences</u> with respect to the baseline of people with eth=Asian.

e.g. if we find $\beta_{\text{Bl}}' > 0$ it's telling us that the average response for people with eth=Bl is higher than that for people with eth=As

Example 2.6.4

c) Does this model suggest there is racial bias in policing actions?

```
some_levels = [k for k in ethnicity_levels if k != 'Asian']
eth_onehot = [np.where(eth==k,1,0) for k in some_levels]

model = sklearn.linear_model.LinearRegression()
model.fit(np.column_stack(eth_onehot), y)
α,βs = model.intercept_, model.coef_

print(f'α = {α}')
for k,β in zip(some_levels, βs):
    print(f'β[{k}] = {β}')
```

This α' parameter is $\hat{P}$(policefindsthg) for eth=Asian

i.e. $\hat{P}$(policefindsthg) is 0.00082 larger for eth=Black than for eth=Asian.

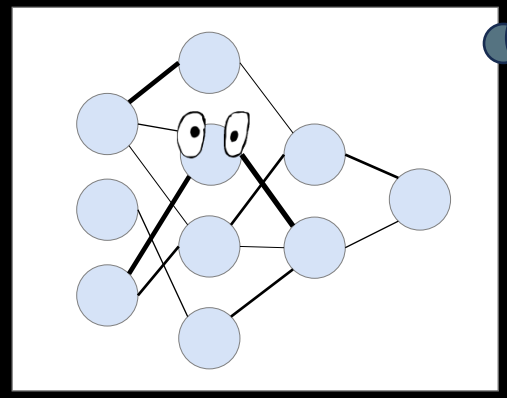
```
α' = 0.261423626892284
β'[Black] = 0.0008154705731562
β'[Mixed] = 0.02871561721115
β'[Other] = -0.0044710573665
β'[White] = -0.003720378247082
```

Are these meaningful?
Is this evidence of bias, or is it just noise?
[See next two weeks]

In a dataset of police stop-and-search records, we've looked for evidence of ethnic bias. What about gender bias? Is the net bias simply additive, or do these biases *intersect*?

Additive model: $y_i \approx \alpha + \beta_{eth_i} + \gamma_{gender_i}$

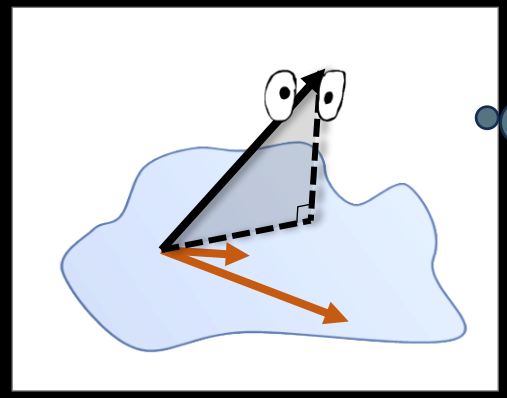
Intersectional model: $y_i \approx \delta_{eth_i, gender_i}$



I can predict exactly what will happen to a person when they're stopped by the police!

Just tell me their gender. And ethnicity. And location. And whether they're left or right handed. And whether they have a pet cat or a dog. And what their pet is called. ...

What was the question again?



Great question! I can answer it directly; all I need is a model with cunningly chosen features.

(just don't ask me whether my model fits well)

(There's a big research area called AI Explainability / Alignment, trying to bridge this gap.)

climate

cl.cam.ac.uk/teaching/2223/DataSci/datasci/ex/climate.html

In [1]:
```
import numpy as np
import pandas
import matplotlib.pyplot as plt
import sklearn.linear_model
π = np.pi
```

Climate dataset challenge

- What is the rate of temperature increase in Cambridge?
- Are temperatures increasing at a constant rate, or has the increase accelerated?
- How do results compare across the whole of the UK?

Your task is to answer these questions using appropriate linear models, and to produce elegant plots to communicate your findings. Please submit a Jupyter notebook, or a pdf. Include explanations of what your models are, and of what your plots show.

The dataset is from https://www.metoffice.gov.uk/pub/data/weather/uk/climate/. Code for retrieving the dataset is given at the bottom.

Upload your answers to Moodle by Sunday evening for presentation / discussion next week