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

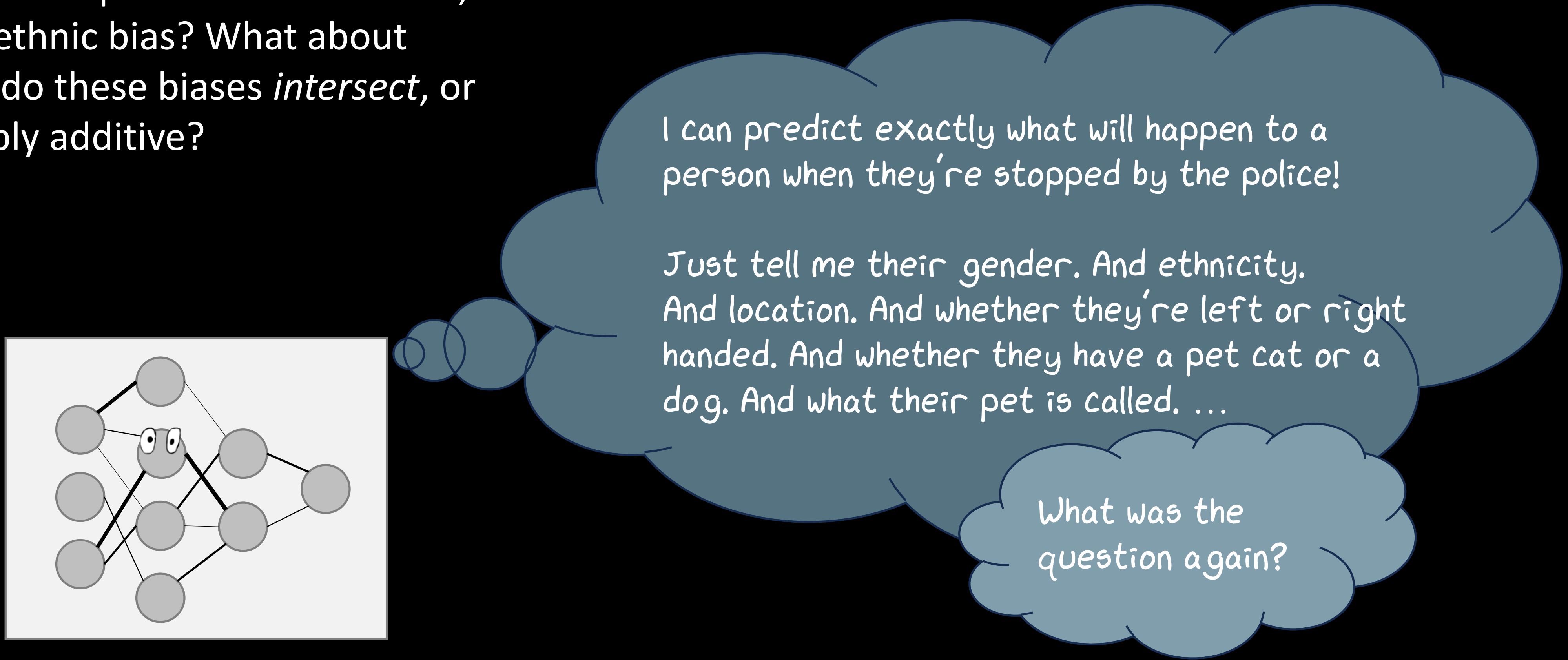
Do I trust this finding?

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

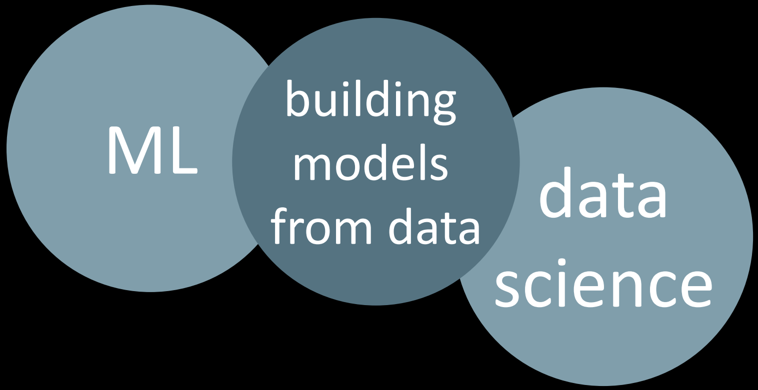
force	Date	LatLng	Object of search	Gender	Age range	Officer-defined ethnicity	Outcome
cambridgeshire	2023-08-31 15:44:04+00:00	(52.43,-0.142)	Controlled drugs				A no further action disposal
cambridgeshire	2023-08-31 15:35:41+00:00	(52.43,-0.142)	Firearms	Male	25-34	White	Khat or Cannabis warning
cambridgeshire	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/

In a dataset of police stop-and-search records, is there evidence of ethnic bias? What about gender bias? If so, do these biases *intersect*, or is the net bias simply additive?



ML mindset
We just want to make good predictions, we don't care about the parameters



Science mindset
We have questions in mind, and we can answer them by looking at our model's parameters

Sign in

The Guardian

News website of the year

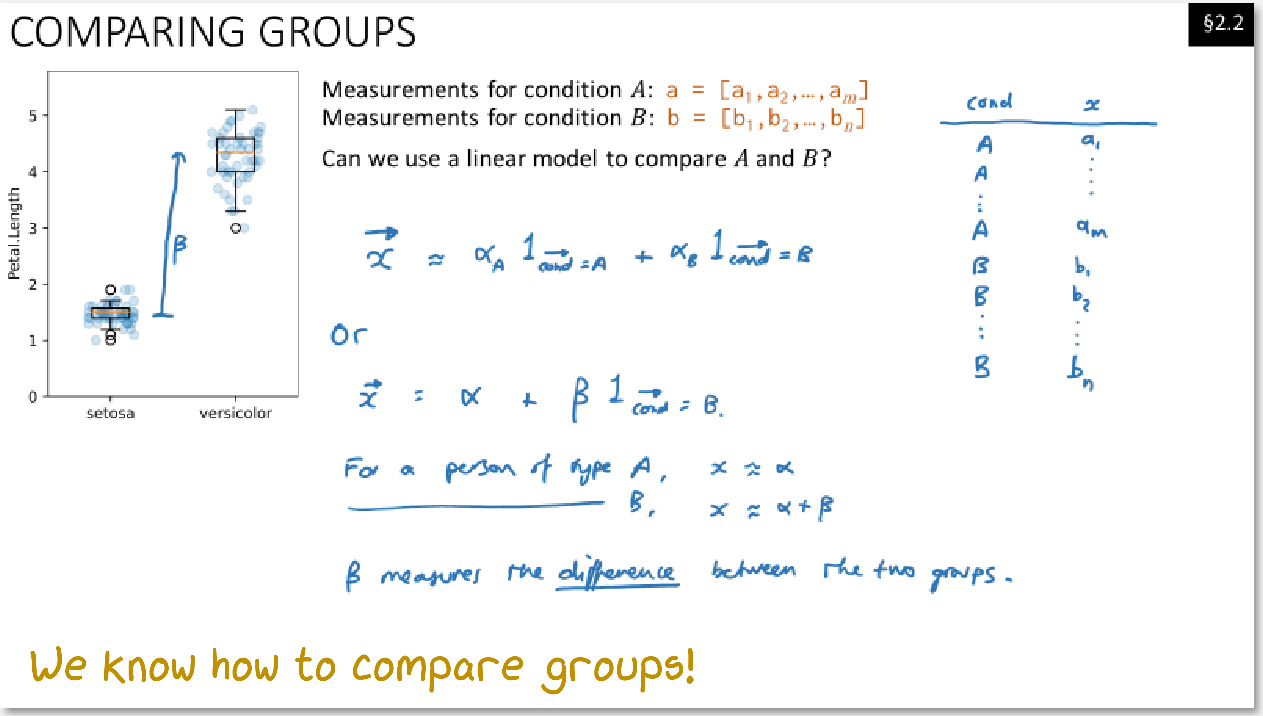
Stop and search

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?



Let y_i be the response in row i, $\;y_i = \begin{cases} 1 & \text{if the police found something} \\ 0 & \text{if the police found nothing} \end{cases}$

The average response in ethnic group k is

$$\begin{matrix}\text{avg } y \\ \text{for ethnicity } k\end{matrix} = \frac{\sum_{i:\text{eth}_i=k} y_i}{|\{i:\text{eth}_i = k\}|} = \frac{\#\text{finds}}{\#\text{stops}} = \mathbb{P}(\text{find something})$$

Let's fit the model $y_i \approx \alpha + \beta_{\text{eth}_i}$

i.e. $\mathbb{P}(\text{find something}) \approx \alpha + \beta_k$ 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.

Let’s fit a model using officer-defined ethnicity as the predictor,

$$y \approx \alpha + \beta_{\text{eth}}$$

Writing it as a linear model with one-hot coding,

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

```python
ethnicity_levels = np.unique(eth)
eth_onehot = [np.where(eth==k,1,0) for k in ethnicity_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(ethnicity_levels, βs):
    print(f'β[{k}] = {β}')
```

```
α = -34037792910.00365
β[Asian] = 34037792910.26522
β[Black] = 34037792910.265717
β[Mixed] = 34037792910.2939
β[Other] = 34037792910.2604
β[White] = 34037792910.261
```

§2.5 The geometry of linear models

NST Maths A, Michaelmas

1.7.2 Shortest distance of a point from a plane

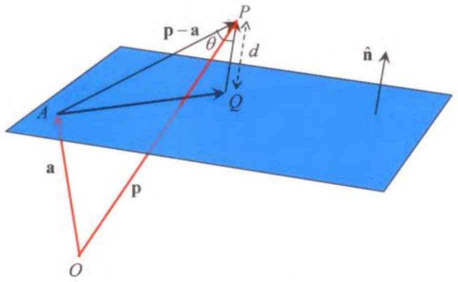


Figure 13: The shortest distance between the point P and the plane defined by the point A and normal $\hat{\mathbf{n}}$.

Consider the plane that passes through point A (given by position vector $\mathbf{a}$) and that has unit normal $\hat{\mathbf{n}}$. From (11), the equation for the plane is defined by the points $\mathbf{r}$ satisfying

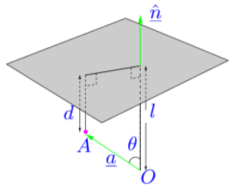
$$(\mathbf{r}-\mathbf{a})\cdot\hat{\mathbf{n}} = 0 .$$

The closest point on the plane to a point P, given by position vector $\mathbf{p}$, is the point Q, where $\overrightarrow{QP}$ is normal to the plane and $|\overrightarrow{QP}| = d$.

NST Maths B, Michaelmas

Example: Distance of point from plane

- What is distance of point A with position vector $\underline{a}$ from plane $\underline{r}\cdot\underline{\hat{n}} = l$?



- Line containing A and point of closest approach of plane to A must be $\parallel \underline{\hat{n}}$; has equation

$$\underline{r} = \underline{a} + \lambda\underline{\hat{n}}$$

- Line meets plane where $\underline{r}\cdot\underline{\hat{n}} = l$, i.e. where

$$l = \underline{a}\cdot\underline{\hat{n}} + \lambda$$

- λ is distance along line from $\underline{a}$ so required distance is $|\underline{a}\cdot\underline{\hat{n}} - l|$

NST Maths A, Easter

Orthogonality - 2/3

- The vectors $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ are a basis of orthonormal vectors in $\mathbb{R}^3$:

$$\mathbf{e}_\mathbf{i}\cdot\mathbf{e}_\mathbf{j} = \delta_{\mathbf{ij}} . \qquad (2)$$

- We use the orthogonality properties (2) to calculate the components of $\mathbf{a}$:

$$\mathbf{e}_1\cdot\mathbf{a} = a_1\times 1 + a_2\times 0 + a_3\times 0 = a_1 .$$

In general

$$a_i = \mathbf{e}_i\cdot\mathbf{a} , \quad \text{for} \quad i = 1,2,3,. \qquad (3)$$

- The above generalises to Euclidean space $\mathbb{R}^n$ with

$$\mathbf{a} = a_i\,\mathbf{e}_i , \quad \text{for} \quad i = 1,\ldots,n,.$$

the components a_i are evaluated in the same way as in the case with $n=3$ because (2) and (3) still hold, but with i, j now in the range 1 to n.

NST Maths B, Easter

Definition. V is called a **vector space over** K, and the elements of V are called **vectors**, if the following **axioms** hold:

A1 For any vectors $\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}\in V$, $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w} = \boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$. **(Associativity.)**

A2 For any vectors $\boldsymbol{u},\boldsymbol{v}\in V$, $\boldsymbol{u}+\boldsymbol{v} = \boldsymbol{v}+\boldsymbol{u}$. **(Commutativity.)**

A3 There is a vector in V denoted $\mathbf{0}$, called the **zero vector** for which $\boldsymbol{u}+\mathbf{0} = \boldsymbol{u}$ $\forall\,\boldsymbol{u}\in V$.

A4 For each vector $\boldsymbol{u}\in V$ there is a vector in V denoted $-\boldsymbol{u}$ for which $\boldsymbol{u}+(-\boldsymbol{u}) = \mathbf{0}$. **(Inverse.)**

A5 For any $a\in K$ and any $\boldsymbol{u},\boldsymbol{v}\in V$, $a(\boldsymbol{u}+\boldsymbol{v}) = a\boldsymbol{u}+a\boldsymbol{v}$.

A6 For any $a,b\in K$ and any $\boldsymbol{u}\in V$, $(a+b)\boldsymbol{u} = a\boldsymbol{u}+b\boldsymbol{u}$.

A7 For any $a,b\in K$ and any $\boldsymbol{u}\in V$, $(ab)\boldsymbol{u} = a(b\boldsymbol{u})$.

A8 For the unit scalar $1\in K$ and any $\boldsymbol{u}\in V$, $1\boldsymbol{u} = \boldsymbol{u}$.

The *subspace spanned* by a collection of vectors $\{e_1, \dots, e_K\}$
is the set of all linear combinations

$$\mathcal{S} = \{\lambda_1 e_1 + \cdots + \lambda_K e_K : \lambda_k \in \mathbb{R} \text{ for all } k\}$$

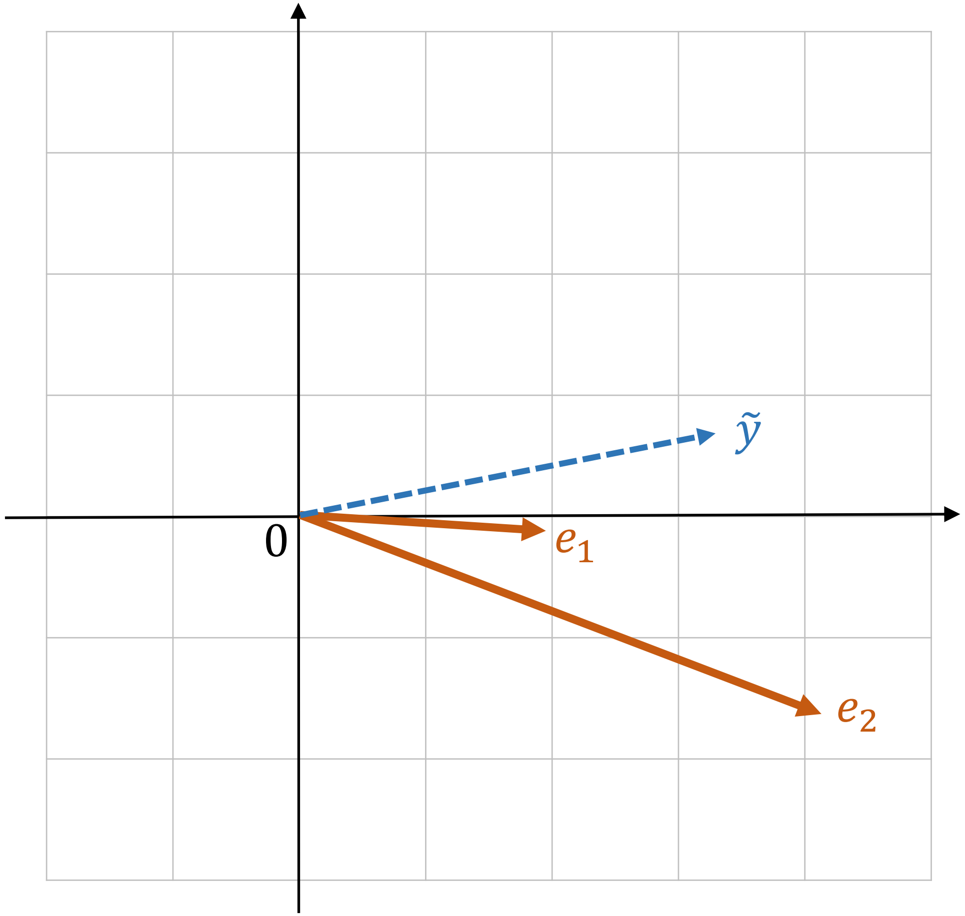
The vectors are *linearly dependent*
if at least one of the e_k can be written as a linear combination of the others,
i.e. there is some set of real numbers $(\lambda_1, \dots, \lambda_K)$ not all equal to zero such that

$$\lambda_1 e_1 + \cdots + \lambda_K e_K = 0$$

If not, they are *linearly independent*, and

$$\lambda_1 e_1 + \cdots + \lambda_K e_K = 0 \quad \Rightarrow \quad \lambda_1 = \cdots = \lambda_K = 0$$

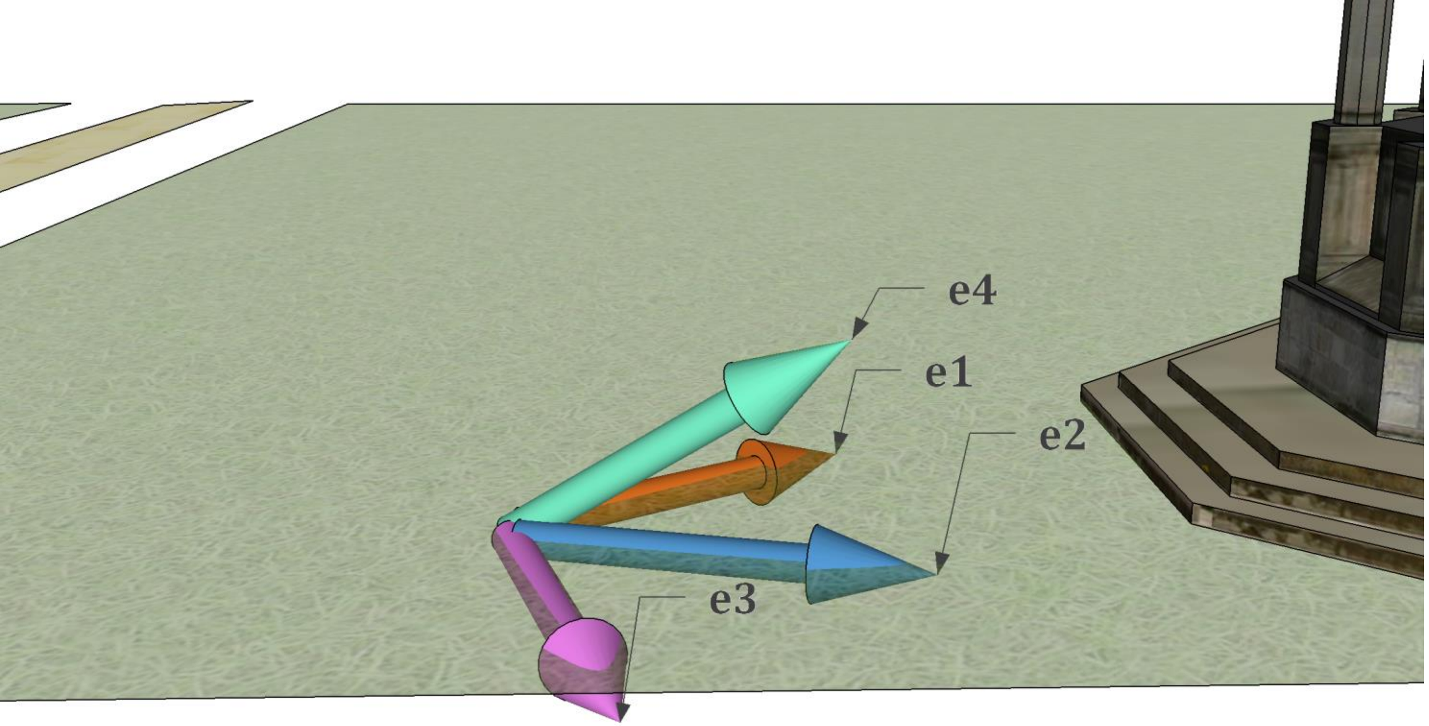
`np.linalg.matrix_rank(np.column_stack(`$[e_1, \dots, e_K]$`))` is $< K$ if linearly dependent
$= K$ if linearly independent



The subspace spanned by $\{e_1, e_2\}$ is $\mathbb{R}^2$

Any $\tilde{y} \in \mathbb{R}^2$ can be written as a linear combination of e_1 and e_2

- by eye, $\tilde{y} = 2.5e_1 - 0.3e_2$



The subspace spanned by $\{e_1, e_2, e_3, e_4\}$ is $\mathbb{R}^3$

Are $\{e_1, e_2, e_3, e_4\}$ linearly independent? No.

If we discarded e_2 …

Are $\{e_1, e_3, e_4\}$ linearly independent? What's the span? They are linearly ind, span is $\mathbb{R}^3$

If we discarded e_1 …

Are $\{e_2, e_3, e_4\}$ linearly independent? What's the span? They are linearly ind, span is $\mathbb{R}^3$.

Exercise 2.5.2

Are the following five vectors linearly independent? If not, find a subset that is.

$$e_1 = [1,1,1,1]$$
$$e_2 = [0,1,1,0]$$
$$e_3 = [1,0,0,1]$$
$$e_4 = [1,1,1,0]$$
$$e_5 = [0,0,0,1]$$

$$e_1 = e_2 + e_3$$
$$e_1 = e_4 + e_5$$

So $\operatorname{span}(\{e_1, e_2, e_3, e_4, e_5\})$

$= \operatorname{span}(\{e_1, e_2, e_3, e_4\})$ since $e_5 = e_1 - e_4$

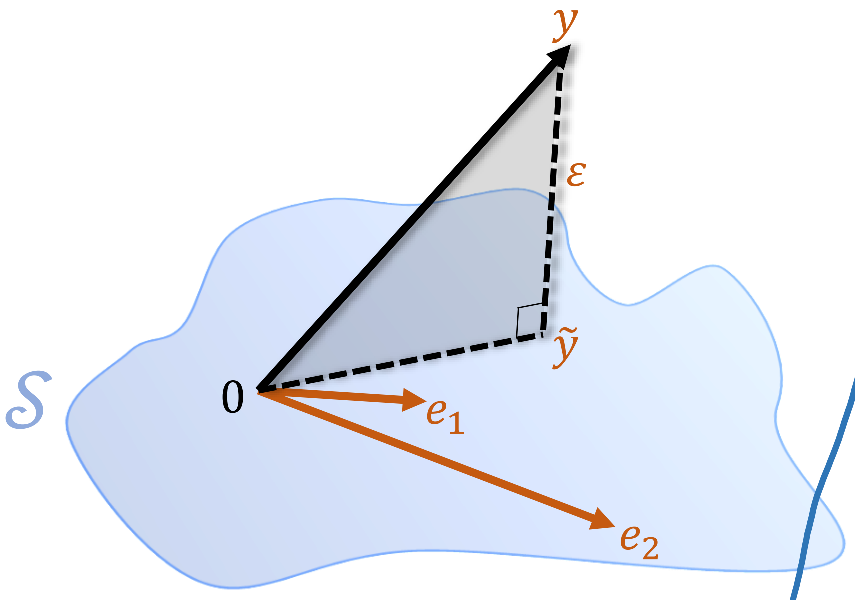
$= \operatorname{span}(\{e_1, e_2, e_4\})$ since $e_3 = e_1 - e_2$.

Q. Are $\{e_1, e_2, e_4\}$ linearly independent?

Suppose $\lambda_1 e_1 + \lambda_2 e_2 + \lambda_4 e_4 = 0$

$$\Rightarrow \quad \lambda_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \lambda_2 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + \lambda_4 \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \quad \left.\begin{array}{l} \lambda_1 + \lambda_4 = 0 \\ \lambda_1 + \lambda_2 + \lambda_4 = 0 \\ \lambda_1 + \lambda_2 + \lambda_4 = 0 \\ \lambda_1 = 0 \end{array}\right\} \Rightarrow \begin{array}{l} \lambda_1 = 0 \\ \lambda_4 = 0 \\ \lambda_2 = 0 \end{array} \Rightarrow$$ they are linearly independent.



What does "closest" mean?

It means: find $\vec{\tilde{y}} \in S$ to minimize $\| \vec{y} - \vec{\tilde{y}} \|$

ie to minimize $\sqrt{\sum_i \varepsilon_i^2}$ where $\vec{\varepsilon} = \vec{y} - \vec{\tilde{y}}$

The minimization is over all $\tilde{y} \in S$.

Every $\tilde{y} \in S$ can be written as a linear combination of $e_1 \cdots e_K$.

What does "best approx." mean?

It means: find $\beta_1 \cdots \beta_K$ to minimize

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

$$\vec{\varepsilon} = \vec{y} - (\beta_1 \vec{e}_1 + \cdots + \beta_K \vec{e}_K)$$

GEOMETRY

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

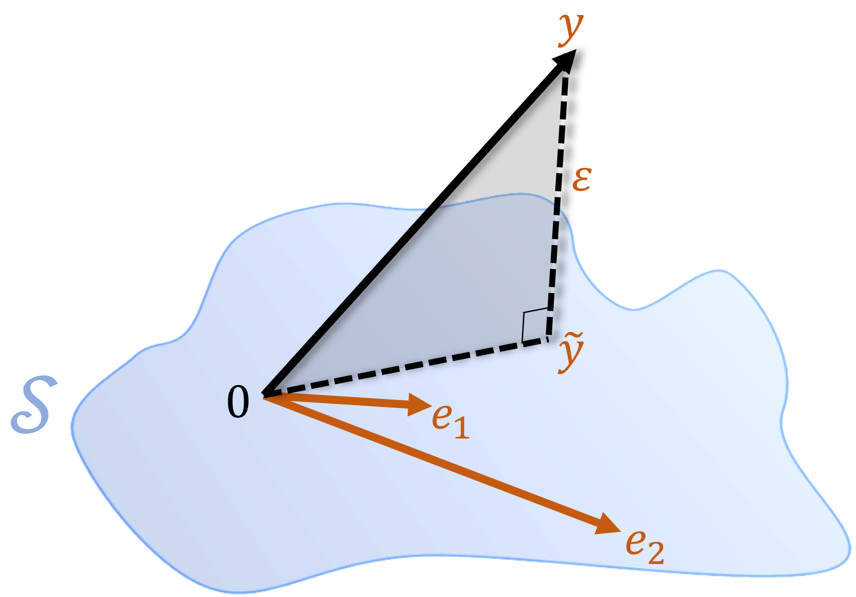
What's the closest we can get to y, while staying in $\mathcal{S}$?

LINEAR MODELLING / LEAST SQUARES ESTIMATION

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

What parameters give us the best approximation?

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



Geometric intuition can help us understand what Least Squares Estimation is doing …

GEOMETRY

Q. Is there a unique way to write $\tilde{y}$ as a linear combination of $\{e_1, \dots, e_K\}$?

A. If they are linearly independent, yes. If they're not linearly independent, maybe not.

LINEAR MODELLING / LEAST SQUARES ESTIMATION

Q. When we run least squares estimation, does it always return the same parameter estimates?

A. If the features are NOT linearly independent, different runs might give different parameters

(and if some other data scientist reports different parameter estimates, our audience will be confused!)

§2.6 Interpreting parameters

- Write out the predicted response for a few typical / representative datapoints.
  This helps see what the parameters mean.

- Write out the features.
  If two models have different features but the same feature space, then (once fitted) they make the same predictions on the dataset.

- Check if the features are linearly dependent.
  If so, the parameters have no intrinsic meaning.
  We say the features are *confounded*, and the parameters are *non-identifiable.*

These three models yielded very different estimates for α. Why?

Model 0: $\texttt{temp} \approx \alpha\,1 + \beta_1 \sin(2\pi\,\texttt{t}) + \beta_2 \cos(2\pi\,\texttt{t})$ $\Rightarrow$ $\hat{\alpha} = 10.6$ °C

Model A: $\texttt{temp} \approx \alpha\,1 + \beta_1 \sin(2\pi\,\texttt{t}) + \beta_2 \cos(2\pi\,\texttt{t}) + \gamma\,\texttt{t}$ $\Rightarrow$ $\hat{\alpha} = -60.2$ °C

Model B: $\texttt{temp} \approx \alpha\,1 + \beta_1 \sin(2\pi\,\texttt{t}) + \beta_2 \cos(2\pi\,\texttt{t}) + \gamma\,(\texttt{t-2000})$ $\Rightarrow$ $\hat{\alpha} = 10.5$ °C

EXERCISE. Write out the predicted response for a few typical / representative datapoints.

Model 0: at $t=0$, pred. temp $= \alpha + \beta_2$ $\longrightarrow$ over a single sinusoid, α is avg. temperature.

Model A: at $t=0$, pred. temp. $= \alpha + \beta_2$ $\longrightarrow$ over a single sinusoid in 1 BC, α is avg. temperature.

$t=2000$, $= \alpha + \beta_2 + 2000\gamma$

Model B: at $t=0$, pred temp $= \alpha + \beta_2 - 2000\gamma$

at $t=2000$, $= \alpha + \beta_2$ $\longrightarrow$ over a single sinusoid in 2000 AD, α is avg. temperature.

These three models yielded very different estimates for α. Why?

Model 0: $\texttt{temp} \approx \alpha\,\texttt{1} + \beta_1 \sin(2\pi\,\texttt{t}) + \beta_2 \cos(2\pi\,\texttt{t})$ $\Rightarrow$ $\hat{\alpha} = 10.6\,°\mathrm{C}$

Model A: $\texttt{temp} \approx \alpha\,\texttt{1} + \beta_1 \sin(2\pi\,\texttt{t}) + \beta_2 \cos(2\pi\,\texttt{t}) + \gamma\,\texttt{t}$ $\Rightarrow$ $\hat{\alpha} = -60.2\,°\mathrm{C}$

Model B: $\texttt{temp} \approx \alpha\,\texttt{1} + \beta_1 \sin(2\pi\,\texttt{t}) + \beta_2 \cos(2\pi\,\texttt{t}) + \gamma\,(\texttt{t-2000})$ $\Rightarrow$ $\hat{\alpha} = 10.5\,°\mathrm{C}$

EXERCISE. Write out the features.

Model A: features $\{\vec{1}, \sin(2\pi\vec{t}), \cos(2\pi\vec{t}), \vec{t}\}$

Model B: features $\{\vec{1}, \sin(2\pi\vec{t}), \cos(2\pi\vec{t}), \underbrace{\vec{t} - 2000\times\vec{1}}\}$.

This can be written as a linear combination of $\vec{1}$ and $\vec{t}$

Model A and B have the same feature space.

They've essentially the same model, just with different parameterization.

Model A asks: tell me the avg. temp in 1BC

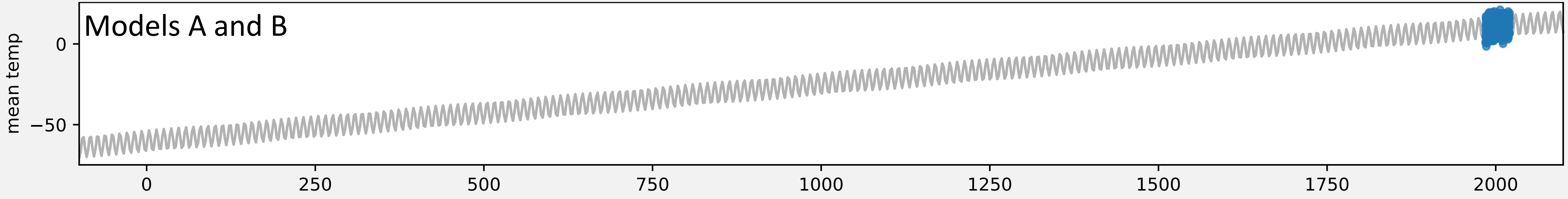
Model B asks: tell me the avg. temp in 2000 AD

These three models yielded very different estimates for α. Why?

Model 0: $\texttt{temp} \approx \alpha\,\texttt{1} + \beta_1 \sin(2\pi\,\texttt{t}) + \beta_2 \cos(2\pi\,\texttt{t})$ $\Rightarrow$ $\hat{\alpha} = 10.6$ °C

Model A: $\texttt{temp} \approx \alpha\,\texttt{1} + \beta_1 \sin(2\pi\,\texttt{t}) + \beta_2 \cos(2\pi\,\texttt{t}) + \gamma\,\texttt{t}$ $\Rightarrow$ $\hat{\alpha} = -60.2$ °C

Model B: $\texttt{temp} \approx \alpha\,\texttt{1} + \beta_1 \sin(2\pi\,\texttt{t}) + \beta_2 \cos(2\pi\,\texttt{t}) + \gamma\,(\texttt{t-2000})$ $\Rightarrow$ $\hat{\alpha} = 10.5$ °C



Models A and B are essentially the same, because they have the same feature space.

But because they use different representations of the feature space,
they report different readouts.

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?

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 equation 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?

(a)

$$y \approx \alpha 1 + \beta_{As} e_{As} + \beta_{Bl} e_{Bl} + \beta_{Mi} e_{Mi} + \beta_{Oth} e_{Oth} + \beta_{Wh} e_{Wh} \quad \text{where} \quad e_k = 1_{\mathrm{eth}=k}$$

ethnic groups: Asian, Black, Mixed, Other, White

(b) These are linearly dependent: $1 = e_{As} + e_{Bl} + e_{Mi} + e_{Oth} + e_{Wh}$

So the parameters are not identifiable, ie we're likely to get silly answers out of linear_model fitting.

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 for presentation / discussion next week