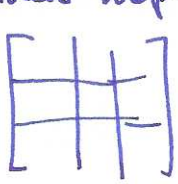
Jordan canonical form. Let $T: V \to V$ be a linear map. Then up to change of basis T is given by a block diagonal matrix:  where each block looks like

$$\begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & 1 & \\ & & \lambda_i & 0 \\ & & & \lambda_j \end{bmatrix} \quad \text{(over } \mathbb{C}\text{!)}.$$

Examples $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ $\begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ $\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. $\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ $\begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$.

A matrix is <u>diagonalizable</u> if its Jordan canonical form is diagonal.

"Generic" matrices have distinct eigenvalues and are diagonalizable.

<u>Observation</u> $A = CDC^{-1}$ $D = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}$

then $A^n = (CDC^{-1})^n = CDC^{-1}CDC^{-1}\cdots CDC^{-1} = CD^nC^{-1} = C\begin{bmatrix} \lambda_1^n & & \\ & \ddots & \\ & & \lambda_d^n \end{bmatrix}C^{-1}$

so if λ_1 is eigenvalue with largest absolute value, then

$$\|A^n\| \sim |A_1|^n$$

<u>Fact/Corollary</u> $\det(A)$ = product of eigenvalues $\lambda_1\lambda_2\cdots\lambda_d$.

<u>Q:</u> how to find $\det(A)$ / eigenvalues? A: row reduction. (no scaling of rows!).

$\uparrow$ $O(d^3)$.

<u>Useful fact</u> If A is a symmetric matrix, then A is diagonalizable.

<u>Proof</u> (special case, eigenvalues all distinct).

Let $v_1, \ldots v_d$ be eigenvectors for A, with corresponding eigenvalues $\lambda_1, \ldots \lambda_d$

Let $C = [v_1\ v_2 \ldots v_d]$ ← matrix of eigenvalues.

<u>claim</u> $C^{-1}AC$ is diagonal matrix $\begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix}$.

Let $e_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix}$ $e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} \ldots$ $e_d = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$.

check $Be_i = \lambda_i e_i$ for each i.

$$C^{-1}ACe_i = C^{-1}Av_i = C^{-1}\lambda_i v_i = \lambda_i C^{-1}v_i = \lambda_i e_i \quad \checkmark \quad \square.$$

Dual vector spaces

Let V be a vector space, the dual vector space or dual space is

$$V^* = \{\text{linear maps from } V \text{ to } \mathbb{R}\}$$
$$= \{v^* : V \to \mathbb{R} \mid v^* \text{ is linear}\}$$

fact V^* is a vector space: $0 : V \to \mathbb{R}$, $v \mapsto 0$.

operations:

$$(a^* + b^*)(v) = a^*(v) + b^*(v)$$
$$(\lambda a^*)(v) = \lambda a^*(v).$$

check linear maps: $a^*(v+w) + b^*(v+w) = a^*(v) + b^*(v) + a^*(w) + b^*(w)$. etc.

Note $T : V \to W$ a linear map defines a dual map $T^* : W^* \to V^*$.

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ \downarrow & & \downarrow \\ \mathbb{R} & & \mathbb{R} \end{array} \qquad T^*(w^*) \in V^*.$$

check T^* is linear and unique.

$$T^*(w^*)(v) = w^*(T(v)).$$

Basic statistical models

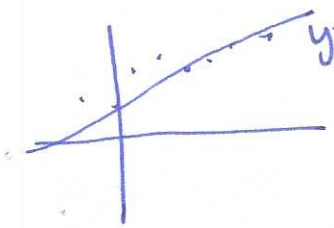
parametric vs non-parametric.

examples: linear regression, kNN

properties:

- fast, get error estimates, only works on linear data.
- don't need to know data is linear, don't get error estimates, need to keep all data.

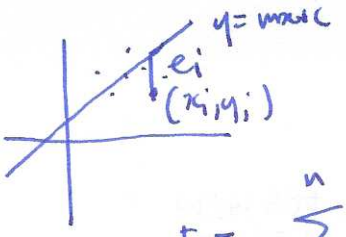
Review of basic linear regression



data: $(x_1, y_1), \ldots, (x_n, y_n)$ ← n points

aim: minimise least squares error LSE

residual sum of squares RSS



$(\text{error})^2 = e_i^2 = (y_i - mx_i - c)^2$.

so $E = \sum_{i=1}^{n} (y_i - mx_i - c)^2 = \sum_{i=1}^{n} (y_i^2 + m^2 x_i^2 + c^2 - 2y_i mx_i - 2y_i c + 2mcx_i)$

Q: what are best choices of m, c? (stats people like: $\hat{\beta}_1, \hat{\beta}_0$).

1. find $\frac{\partial E}{\partial c} = \sum_{i=1}^{n} (2c - 2y_i + 2mx_i)$ and solve $\frac{\partial E}{\partial c} = 0$:

$\sum_{i=1}^{n} (2c - 2y_i + 2mx_i) = 0 \qquad 2cn - 2\sum y_i + 2m\sum x_i = 0$

$c = \frac{1}{n}\sum y_i - \frac{1}{n} m \sum x_i = \bar{y} - m\bar{x}$

($\bar{y}$: average of y_i's; $\bar{x}$: average of x_i's.)

2. find $\frac{\partial E}{\partial m} = \sum_{i=1}^{n} (2mx_i^2 - 2y_i x_i + 2cx_i)$, with $c = \bar{y} - m\bar{x}$

$= \sum_{i=1}^{n} (2mx_i^2 - 2y_i x_i + 2\bar{y}x_i - 2m\bar{x}x_i)$. solve $\frac{\partial E}{\partial m} = 0$.

$m\left((\sum x_i^2) - \bar{x}\sum x_i\right) = \sum x_i y_i - \bar{y}\sum x_i$

$m = \dfrac{\left(\sum_{i=1}^{n} x_i y_i\right) - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} \overset{\text{fact}}{=} \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$. ⊛

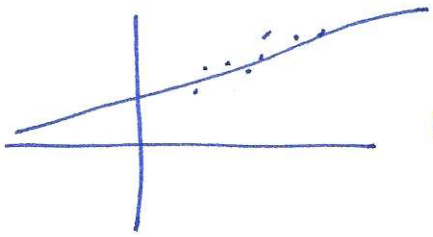
check ⊛: $\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} (x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\bar{x}^2 n + n\bar{x}^2$

$= \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$

$$\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) = \sum(x_iy_i - \bar{x}y_i - \bar{y}x_i + \bar{x}\bar{y})$$

$$= \sum x_iy_i - n\bar{x}\bar{y} - n\bar{y}\bar{x} + n\bar{x}\bar{y} = \sum x_iy_i - n\bar{x}\bar{y} \quad \square.$$

Summary



$(x_i,y_i)_{i=1}^n$, best fit line is $y = mx+c$

with $$m = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} = \frac{(\sum x_iy_i) - n\bar{x}\bar{y}}{(\sum x_i^2) - n\bar{x}^2}$$

and $$c = \bar{y} - m\bar{x}.$$

note m, c depend on: $\sum x_i$, $\sum y_i$, $\sum x_iy_i$, $\sum x_i^2$ } don't need to remember all the data.

Error estimates (assuming data is linear; errors, indep of i, Gaussian)...

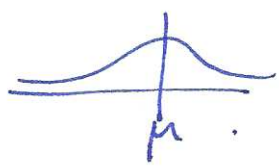
$$SE(c)^2 = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right] \qquad SE(m)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}.$$

where $\sigma^2 = \mathrm{Var}(\epsilon)$ ← variance of errors (approx by $\sum \epsilon_i^2 - \bar{\epsilon}^2$.) (in fact divide by $(n-2)$).

so can get confidence intervals:

95% confidence intervals are $m \pm 2\,SE(m)$, $c \pm 2SE(c)$. } if these values very large, probably not a good model.

Recall: sample n points from a normal dist $N(\mu,\sigma^2)$. $x_1,\dots,x_n$



$\bar{x} = \frac{1}{n}\sum x_i$ sample mean has dist $N(\mu, \frac{\sigma^2}{n})$.

Problem don't know σ in $N(\mu,\sigma^2)$. **Solution**: use sample variance / standard deviation approx σ by $\sigma(x_1,\dots,x_n)$.

(should use Student t-dist....).