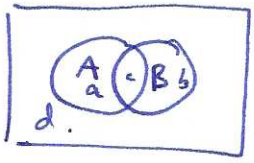
Bayes Thm

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

$$= \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$



$a+b+c+d = 1$

$\mathbb{P}(A) = a+c$

$\mathbb{P}(B) = c+b.$

$$\mathbb{P}(A|B) = \frac{c}{c+b}.$$

$\mathbb{P}(A \cap B) = c$

$$\mathbb{P}(B|A) = \frac{c}{a+c}$$

so $\mathbb{P}(B|A).\mathbb{P}(A) = c.$

Example. you take a covid test.

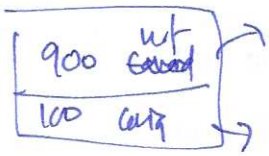
- positive
- negative

} what is the probability you have covid?

your test is 95% accurate, i.e. if you ~~test positive, you have~~ are positive, test is positive 95% of the time.
are negative next 95% of the time.

your 100% of the population have covid.

1000 people ⇝ 100 have covid.
⇝ 900 don't.



~~$\mathbb{P}$(test +ve | have covid)~~

$$\mathbb{P}(\text{have covid} \mid \text{test positive}) = \frac{\mathbb{P}(\text{have covid} \cap +ve)}{\mathbb{P}(+ve)} = \frac{95/1000}{130/1000} \approx 0.73\%.$$

Q: your 1% of pop has covid?

The Bayes classifier

Data set $X = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\}$ $\quad x_i \in \mathbb{R}^d.$

$y_n \in \{1, \dots, n\} = Y.$ ↑ classification buckets.

classifying function $\hat{f} : \mathbb{R}^d \to Y.$ $\quad \hat{y}_i = \hat{f}(x_i).$

error rate: $\frac{1}{n} \sum_{i=1}^{n} \underbrace{I(y_i \neq \hat{y}_i)}_{= \#\text{errors}}$ ← indicator function $I(x,y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$

$= \frac{\#\text{errors}}{n}.$

test error rate: error rate over test data. (having found $\hat{f}$ from training data).

Thm/Fact: the error rate is minimized by the classifier that assigns each observation to the most likely class given its predictor values, i.e.

for which $\mathbb{P}(Y = j \mid X = x_i)$ is largest.

special case $Y = \{0, 1\}$. $\mathbb{P}(Y=0 \mid X=x_i) > \frac{1}{2}$ choose 0

$< \frac{1}{2}$ 1.

Bayes error rate is $1 - \max_j \mathbb{P}(Y=j \mid X=x_i)$.

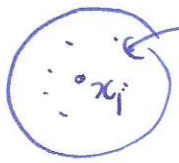
overall Bayes error rate is: $1 - \mathbb{E}\left(\max_j \mathbb{P}(Y=j \mid X)\right)$.

Problem for real data, we do not know the conditional distribution $\mathbb{P}(Y \mid X)$ ∴.

Work around: • make heroic assumptions about prob. dist. of data

• estimate $\mathbb{P}(Y \mid X)$ from data...

↳ this is exactly what KNN does.



find K nearest neighbors

choose max class.