Reminders

↳ Notes (investing time is worth it)

↳ Office Hours: M + W 12:30 - 2,
and by appointment

↳ Discord
- DM
- Pset Questions

Context

Regression : $\underline{x}^{(i)} \in \mathbb{R}^d$ $\quad y^{(i)} \in \mathbb{R}$

Classification: $\underline{x}^{(i)} \in \mathbb{R}^d$ $\quad y^{(i)} \in \{0, \ldots, k\}$

↳ spam or not spam (binary)

↳ objects in an image

↳ next word

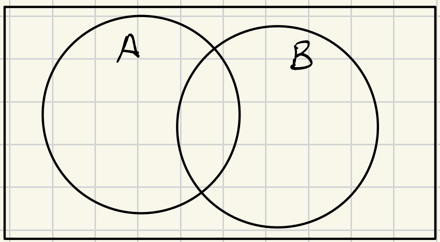
Today: Probability background

Probability

A, B random events

it rains tomorrow

I carry an umbrella



$Pr(A)$

$Pr(A \cup B)$

$Pr(A \cap B)$

$Pr(A \mid B)$

Properties:

$$0 \leq Pr(A) \leq 1$$

$$Pr(A \cap B) = Pr(A)\,Pr(B \mid A)$$
$$= Pr(B)\,Pr(A \mid B)$$

$$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$$

"complement"

$\neg A$ = not A

$$Pr(\neg A) + Pr(A) = 1 \iff Pr(\neg A) = 1 - Pr(A)$$

A, B indep iff $Pr(A \cap B) = Pr(A)\,Pr(B)$

$\boxed{\text{Random Variables:}}$

$X \sim D$

$\sim$ sampled from

D a distribution

For example, $X \sim$ Dice roll

$Pr(X=1) = Pr(X=2) = \ldots = 1/6$

$\boxed{\text{Bayes Rule:}}$

$$Pr(A|B) = \frac{Pr(B|A)\, Pr(A)}{Pr(B)}$$

Proof?

Maximum A Posteriori

Our regression strategy: Choose model that most likely generated data

Classification analogy: Choose label that is most likely given data

Example: $\underline{X}$ = email , $y = \mathbb{1}[\text{email is spam}]$

Posterior :

$$\Pr(Y=1 \mid X=x) \quad \text{vs.} \quad \Pr(Y=0 \mid X=x)$$

Compute using Bayes Rule!

$$\Pr(Y=1 \mid X=x) = \frac{\Pr(Y=1)\,\Pr(X=x \mid Y=1)}{\Pr(X=x)} = \frac{\text{prior} \quad \text{likelihood}}{\text{evidence}}$$

$\boxed{\text{Medical Example:}}$

$X \in \{0,1\}$ outcome of test

$Y \in \{0,1\}$ disease

- Rare (1 % population)
- 5 % FPR
- 10 % FNR

Naive Bayes Classifier (with binary features)

assume independent features

Naive assumption : $$\Pr(X=x \mid Y=1) = \prod_{i=1}^{d} \underbrace{\Pr(X_i = x_i \mid Y=1)}_{p_i^{(1)} \text{ or } (1-p_i^{(1)})}$$

For example, $\Pr\left(X = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \middle| Y=1 \right) = (1-p_1)\, p_2 (1-p_3)(1-p_4)\, p_5$

1. Compute priors $\Pr(Y=1),\ \Pr(Y=0)$

2. Compute observed probs $p_i^{(1)} = \Pr(X_i = 1 \mid Y=1),\ p_i^{(0)} = \Pr(X_i = 1 \mid Y=0)$

3. Compute likelihoods

$$\Pr(X=x \mid Y=1) = \prod_{i=1}^{d} \Pr(X_i = x_i \mid Y=1)$$

4. Predict class with higher posterior $$\operatorname*{argmax}_{i \in \{0,1\}} \frac{\Pr(X=x \mid Y=i)\,\Pr(Y=i)}{\Pr(X=x)}$$