Welcome to Week 2 of Online ML

Reminders: Gradiance 5 – Due Tomorrow

PA2 – Due on April 13th

Probabilistic/Statistical Methods:

Categorical Random Variables

$\mathcal{X} = \{H, T\}$	$\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$
pmf: $p(X = H)$ $p(X = T)$	$p(X=1)$ $p(X=2)$ $\vdots$ $p(X=6)$
$p(X=T) = 1 - p(X=H)$	$p(X=6) = 1 - \sum_{i=1}^{5} p(X=i)$

If a random variable takes K value

$$|\mathcal{X}| = K$$

Then we need (K−1) probabilities to completely specify its pmf.

These are the parameters of ~~ou~~ the distribution

How much data do you need?

e.g. a coin toss

$D = \{H\}$ ✗

$D = \{H H H T T H H T H H\}$ ✓

If we have a roll of a dice would need more data.

$X = \{$ roll of a dice, toss of a coin $\}$

$\mathcal{X} =$ 1, H
1, T
2, H
2, T
⋮

$|\mathcal{X}| = 12$

$\#$ params $= 12 - 1 = 11$

Tumor example:

Shape	Size	Color	smoothness	thickness
cir	large	dark	smooth	thick
Oval	Small	light	irregular	thin

Target malignant (1) $\}$ Y
benign (0)

Y - random variable (Bernoulli)

$P(Y = \text{benign} \mid X = x^*)$

$P(Y = \text{malignant} \mid X = x^*)$

$$\boxed{Y \sim Ber(\theta)}$$

$P(Y = \text{malignant}) = \theta$

$P(Y = \text{benign}) = 1 - \theta$

Given just the labels on the training data.

I can estimate θ:

$$\hat{\theta}_{MLE} = \frac{N_1}{N}$$

$$\hat{\theta}_{MAP} = \frac{N_1 + a - 1}{N + a + b - 2}$$

$P(Y = \text{malignant}) = \theta$

This is not $P(Y = \text{malignant} \mid X = x^*)$

$P(Y = \text{malignant} \mid X = x^*)$

$= P(X = x^* \mid Y = \text{malignant}) \, P(Y = \text{maligna}$

$P(X=x^* | Y=\text{malignant})\, P(Y=\text{mal.})$
$+ P(X=x^* | Y=\text{benign})\, P(Y=\text{benign})$

$P(Y=\text{benign} | X=x^*) = 1 - P(Y=\text{mal.} | X=x^*)$

$P(X=x^* | Y=\text{malignant})$
$P(X=x^* | Y=\text{benign})$
→ Class-conditional probabilities

X is a random variable

multi-variate categorical r.v.

For a 5-D random variable where each feature is binary -

We can have 2^5 possible combinations

$$|X| = 2^5$$

We need to estimate $(2^5 - 1)$ parameters

We need to estimate these parameters for both classes

$P(X=x^* / Y=\text{mal.})$

$P(X=x^* / Y=\text{benign})$

$2 \times (2^5 - 1)$ parameters to estimate

All features are independent of each other

Naive Bayes Classifier

$$\boxed{P(\underline{\cancel{X} = x^*} \mid y = \text{malignant}) = \prod_{j=1}^{D} \underline{P(x_j = x^*_j \mid y = \text{malignant})}}$$

Each $P(\underline{x_j = x^*_j} \mid y = \text{malignant})$ can be modeled as a Bernoulli r.v.

If each feature is binary, then we need $2 * D$ parameters

Training $P(y = \text{mal.})$, $P(y = \text{benign})$ $\gg$ $\theta_{\text{mal.}}$ θ_1, θ_{benign} θ

$P(x_j = 1 \mid y = \text{mal.})$ $P(x_j = 1 \mid y = \text{ben})$
$P(x_j = 0 \mid y = \text{mal})$ $P(x_j = 0 \mid y = \text{benign})$

$\forall j$

$p(X=x, Y=y) = \ldots$

For a training example x_i, y_i

$x_1 \quad y_1$

$x_2 \quad y_2$

$\vdots$

$$P(X=x_i, Y=y_i)$$

$$= P(Y=y_i)\,P(X=x_i \mid Y=y_i)$$

$$ll(D \mid \Theta) = N_1 \log \theta + N_2 \log(1-\theta)$$

$$+ \sum \left(N_{1j} \log \theta_{1j} + (N_1 - N_{1j}) \log(1-\theta_{1j})\right)$$

$$+ \sum - - - - -$$

all parameters in the model

$$\theta = \frac{N_1}{N}$$

N_1 ← # examples with class 1

N ← # examples

$$\boxed{\hat{\theta}_c = \frac{N_c}{N}} \quad \forall c$$

$$\boxed{\hat{\theta}_{cj} = \frac{N_{cj}}{N_c}} \quad \forall j, c$$

These are MLE estimates

How to get a prediction for x^*

$P(Y = mal \mid X = x^*) =$ ______

refer to the ~~Na~~ Bayes rule expression mentioned earlier.

If $P(Y = mal \mid X = x^*) \geq 0.5$ $y^* = mal.$

< 0.5 $y^* = ben$

$$P(Y = mal. \mid X = \{cir., sm, li\})$$

$$= \frac{P(X = \{cir, sm, li\} \mid Y = mal)\, P(Y = mal)}{\left[P(X = \{cir, sm, li\} \mid Y = mal)\, P(Y = mal) + P(X = \{cir, sm, li\} \mid Y = ben)\, P(Y = ben)\right]}$$

$$= P(X_1 = cir \mid Y = mal)\, P(X_2 = sm \mid Y = mal)\, P(X_3 = li \mid Y = mal)$$

$(\qquad)$

$$p(y = mal) = \frac{N_{mal}}{N} = 0.5$$

$$p(y = benign) = \frac{N_{ben}}{N} = 0.5$$

$P(X_1 = cir \mid y = mal) = \frac{3}{5}$	$P(x_1 = cir \mid y = benign) = \frac{[illegible]}{5}$
$P(X_2 = sm \mid y = mal) = \frac{1}{5}$	$P(x_2 = sm \mid y = ben) = \frac{3}{5}$
$p(x_3 = li \mid y = mal) = \frac{2}{5}$	$P(x_3 = li \mid y = ben) = \frac{3}{5}$

$$p(y = mal \mid x = \{cir, sm, li\})$$

$$= \frac{0.5 * \frac{3}{5} * \frac{1}{5} * \frac{2}{5}}{0.5 * \frac{3}{5} * \frac{1}{5} * \frac{2}{5} + 0.5 * \frac{2}{5} * \frac{3}{5} * \frac{3}{5}}$$

$$= \frac{6}{6 + 18} = \underline{0.25}$$

<u>Y is benign</u>

A continuous equivalent of NBC

Quadratic Discriminant Analysis

Linear Discriminant Analysis

QDA / LDA