CS 4342 - Machine Learning

WPI

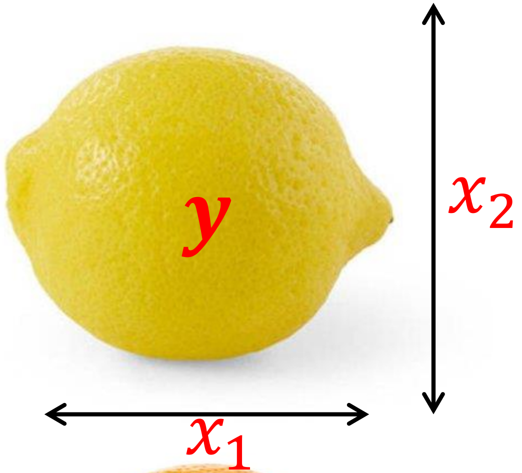
Fall 2019

Lecture 2: *Bayes Rule & Naïve Classifier*
MLE and MAP

Instructor: Ali Yousefi

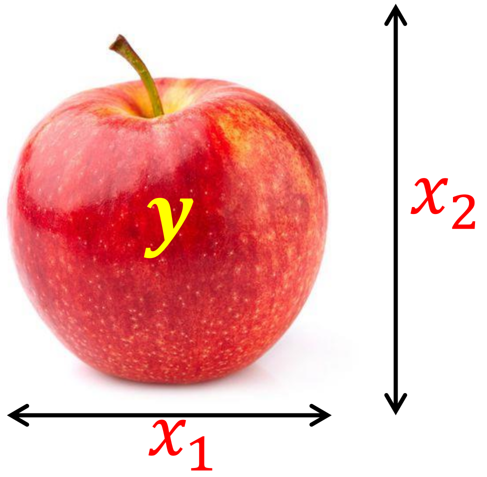
Joint Probability Distribution

$$P(X_1 = x_1, X_2 = x_2, Y = y)$$



Conditional Probability Distribution

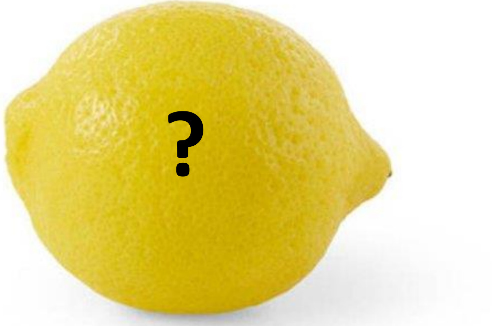
$$P(X_1 = x_1, X_2 = x_2 | Y = y)$$



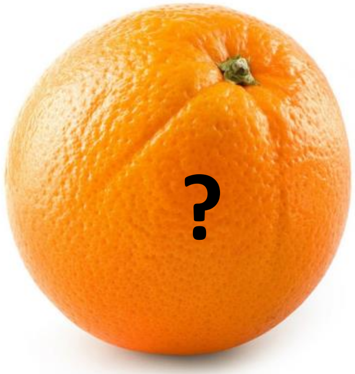
Bayes Theorem

$$P(Y = y | X_1 = x_1, X_2 = x_2)$$

$$= \frac{P(X_1 = x_1, X_2 = x_2 | Y = y) P(Y = y)}{P(X_1 = x_1, X_2 = x_2)}$$



(x_1, x_2)



Bayes Theorem

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

$$P(Y|X) = \frac{P(X|Y)P(Y)}{\sum P(X|Y)P(Y)}$$

Bayes Theorem

- Parametric models

$$P(Y|X,\theta) = \frac{P(X|Y,\theta)P(Y|\theta)}{P(X|\theta)}$$

$$P(Y|X,\theta) = \frac{P(X|Y,\theta)P(Y|\theta)}{\sum P(X|Y,\theta)P(Y|\theta)}$$

Bayes Theorem

Class Conditional Density or Likelihood

Class Prior Probability

$$P(Y|X,\theta) = \frac{P(X|Y,\theta)P(Y,\theta)}{P(X,\theta)}$$

Posterior Probability

Predictor Prior Probability

Model Parameter

We focus on likelihood function

$$L(X;Y,\theta) = P(X|Y,\theta)$$

Log likelihood

$$LL(X;Y,\theta) = logP(X|Y,\theta)$$

Observation

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$$

Basic Classifier

$$P(Y = c|X, \theta\ ) \propto P(X|Y = c, \theta)P(Y = c|\theta)$$

$$\hat{Y} = \operatorname*{argmin}_{c} P(Y = c|X, \theta\ )$$

Basic Distribution

$$\boldsymbol{P(\theta|a,b\,) = \frac{\theta^{a-1}(1-\theta)^{b-1}}{B(a,b)}}$$

$\theta \in [0,1] \qquad a, b > 0$

$$E[\theta] = \frac{a}{a+b} \qquad \mathrm{var}[\theta] = \frac{ab}{(a+b)^2(a+b+1)}$$

$$\mathrm{Mode} = \frac{a-1}{(a+b-2)} \qquad a, b > 1$$

Zero Count Problem

Bayesian (MAP)
You are no good
when sample
size is small

Frequentists (MLE)
You give a different
answer for
different priors

A Basic Classifier: Credit Rating Prediction

- x_1 income level $x_1 = \{'low', 'med', 'high'\}$
- y credit rating – $y = \{'bad', 'good'\}$
- We have a dataset containing 577 samples $D = \{x_i, y_i\}\ \ i = 1, \ldots, 577$
- x is the feature - $x = \{0,1,2\}$

Feature (x)	Credit Rate (y)	
	# bad	#good
X=0	135	105
X=1	165	148
X=2	15	9

- Use Bayes theorem to build a classifier – predict y given x

Class prior probability

Feature (x)	Credit Rate (y) # bad	 #good
X=0	135	105
X=1	165	148
X=2	15	9

Credit Rate (y) # bad	 #good	# bad + #good
315	262	577
0.545	0.455	1
$\text{p}(y = 'bad')$	$\text{p}(y =' good')$	

Class conditional probability

Feature (x)	Credit Rate (y) # bad	#good
x=0	0.428	0.400
x=1	0.524	0.564
x=2	0.048	0.036

$$p(x=?\,|y='bad')$$

$$p(x=?\,|y='good')$$

Posterior probability

$$p(y=?\,|x=1) \propto p(x=1|y=?\,)p(y=?\,)$$

$? = 'good'$ $\quad$ $p(y='\,good'|x=1) \propto 0.564 \times 0.455 = 0.256$

$? = '\,bad'$ $\quad$ $p(y='\,bad'|x=1) \propto 0.524 \times 0.545 = 0.286$

The sum is not equal to 1!

$$p(y='\,bad'|x=1) > p(y='\,good'|x=1)$$

The confident about the decision might not be that strong!

Predictor prior probability

Feature (x)			
X=0	240	0.416	$p(x = 0)$
X=1	313	0.542	$p(x = 1)$
X=2	15	0.042	$p(x = 2)$

$p(y = 'good'|x = 1)$=0.256/0.542=0.472

$p(y =' bad'|x = 1)$=0.528

Posterior probability

$$p(y=?\,|x=1) \propto p(x=1|y=?\,)p(y=?\,)$$

$? = 'good'$ $\quad p(y='\,good'|x=1) \propto 0.564 \times 0.455\text{=}0.256$

$? =' \,bad'$ $\quad p(y='\,bad'|x=1) \propto 0.524 \times 0.545\text{=}0.286$

$$p(y=?\,|x_1=1, x_2=0) \propto p(x_1=1|y=?\,)p(x_2=0|y=?\,)p(y=?\,)$$

Naïve Bayes Classifier

Beta Binomial Model