CSE 575: Statistical Machine Learning

Mid-Term 1

Instructor: Prof. Hanghang Tong
September 12th, 2017

First Name:			
Last Name:			
Email:			
ASU ID:			
Q	Topic	Max Score	Score
1	MLE	20	
2	Continuous Bayes Classifier	20	
3	Discrete Bayes Classifier	20	
4	Naive Bayes Classifier	30	
5	Bayes Classier vs. Naive Bayes Classier	10	
Total:		100	

- This exam book has **11** pages, including this cover page and a blank page at the end.
- Good luck!

1 Maximum Likelihood Estimation (20 points)

Suppose we have a 1-dimensional random variable X, and its pdf $f(X)$ (probability density function) is defined as $f(X) = \frac{1}{b-a-4}$ for $a \leq X \leq 6$ or $10 \leq X \leq b$ and $f(X) = 0$ $otherwise$, where a and b two unknown parameters.

If we draw five data points x_1, x_2, x_3, x_4 and x_5 independently from this distribution, and we observe that $x_1 = 1$, $x_2 = 5$, $x_3 = 10$, $x_4 = 15$ and $x_5 = 12$.

1. [**5 pts.**] What is the likelihood L of observing $\{x_1, x_2, x_3, x_4, x_5\}$? (**Hints:** the likelihood L is function of the two parameters a and b.)

   Solutions: $L = (\frac{1}{b-a-4})^5$ if $a \leq 1$ and $b \geq 15$. Otherwise $L = 0$ (lose two points if not specifying when L is zero; lose two points if the condition for a non-zero L is wrong).

2. [**10 pts.**] What is the maximum likelihood estimation of a (5 points)? What is the maximum likelihood estimation of b (5 points)? Justify your answer.

   Solutions: $a = 1$ and $b = 15$ (we want a as large as possible and b as small as possible, in order to maximize L. in the meanwhile, $a \leq 1$ and $b \geq 15$ in order to have a non-zero L.

3. [**5 pts.**] What is the likelihood of $\{x_1, x_2, x_3, x_4, x_5\}$ given your MLE estimation of a and b?

   Solutions: $L = (0.1)^5$.

2 Continuous Bayes Classifier (20 points)

We want to build a Bayes Classifier for a binary classification task ($y = 1$ or $y = 2$) with a 1-dimensional input feature (x). We know the following quantities: (1) $P(y = 1) = 0.8$; (2) $P(x|y = 1) = 0.5$ for $1 \leq x \leq 2$, $P(x|y = 1) = 0.25$ for $2 < x \leq 4$ and $P(x|y = 1) = 0$ otherwise; and (3) $P(x|y = 2) = 0.5$ for $3 \leq x \leq 5$ and $P(x|y = 2) = 0$ otherwise.

- **[2pts]**. What is the prior of the class label $y = 2$?

  Solutions: $P(y = 2) = 0.2$

- **[3pts]**. What is $P(y = 1|x)$?

  Solutions: $P(y = 1|x) = 1 \ \ (1 \leq x < 3)$, $P(y = 1|x) = 2/3 \ \ (3 \leq x \leq 4)$ and 0 otherwise (1 pt for each case)

- **[3pts]**. What is $P(y = 2|x)$?

  Solutions: $P(y = 2|x) = 1/3 \ \ (3 \leq x \leq 4)$, $P(y = 2|x) = 1 \ \ (4 < x \leq 5)$ and 0 otherwise

- **[4pts]**. For $x = 1$, what is class label your classifier will assign? What is the risk of this decision?

  Solutions: $y = 1$, risk is 0 (2 pts for each)

- **[4pts]**. For $x = 3.5$, what is class label your classifier will assign? What is the risk of this decision?

  Solutions: $y = 1$, risk is $1/3$ (2 pts for each)

- **[2pts]**. What is the decision boundary of your Bayes classifier?

  Solutions: $[1, 4] : \ y = 1$; $(4, 5] : \ y = 2$; tie/unknown otherwise. (1 pt for each segment. lose 0.5pt if mistaken on the boundary)

- **[2pts]**. What is the Bayes error of your Bayes classifier?

  Solutions: 0.1

3 Discrete Bayes Classifier (20 points)

We want to build a Bayes Classifier for a binary classification task ($y = 1$ or $y = 2$) with one discrate feature x, where $x \in \{1, 2, 3, 4, 5, 6\}$. We know the following quantities: (1) $P(y = 1) = 0.5$; (2) $P(x = 1|y = 1) = 0$, $P(x = 2|y = 1) = 0.1$, $P(x = 3|y = 1) = 0.5$, $P(x = 4|y = 1) = 0.2$, $P(x = 5|y = 1) = 0.1$, $P(x = 6|y = 1) = 0.1$; and (3) $P(x = 1|y = 2) = 0.1$, $P(x = 2|y = 2) = 0.2$, $P(x = 3|y = 2) = 0.2$, $P(x = 4|y = 2) = 0.3$, $P(x = 5|y = 2) = 0.2$, $P(x = 6|y = 2) = 0$.

- **[2pts]**. What is the prior of the class label $y = 2$?

  Solutions: $P(y = 2) = 0.5$

- **[4pts].** Put all the $P(x|y = 1)$ and $P(x|y = 2)$ numbers into a 6×2 table, whose rows correspond to six different values of x (i.e., $x = 1, 2, ..., 6$), and two columns correspond to two different values of y (i.e., $y = 1$ and $y = 2$).

  Solution:

0	0.1
0.1	0.2
0.5	0.2
0.2	0.3
0.1	0.2
0.1	0

- **[4pts]**. What is $P(y = 1|x = 3)$? What is $P(y = 2|x = 3)$?

  Solutions: $5/7$ and $2/7$ (2 pts each).

- **[2pts]**. For an example with the following feature $x = 6$, what is class label your classifier will assign? What is the risk of this decision?

  Solutions: $y = 1$, risk is 0 (1 pt for each)

- **[2pts]**. For an example with the following feature $x = 1$, what is class label your classifier will assign? What is the risk of this decision?

  Solutions: $y = 2$, risk is 0 (1 pt for each)

- **[3pts]**. What is the decision boundary of your Bayes classifier?

Solutions: $x = 3, 6$, $y = 1$, otherwise $y = 2$ (each wrong decision for one x value, lose half a point).

- **[3pts]**. What is the Bayes error of your Bayes classifier?

  Solutions: 0.3

4 Naive Bayes Classifier (30 points)

Given the training data set in the following Table, we want to train a binary classifier, with (1) the last column being the class label y (i.e., $y = 1$ or $y = 0$); (2) x_1, x_2 and x_3 being three binary features; and (3) each row being a training data point.

Data	x_1	x_2	x_3	y
1	0	0	0	1
2	0	0	1	0
3	0	1	1	0
4	0	1	1	0
5	0	0	1	1
6	1	0	1	1
7	1	0	1	0
8	1	0	1	0
9	1	1	1	1
10	1	0	1	1

1. **[9 pts.]** How many independent parameters are there in your Naive Bayes classifier? What are they? Justifiy your answer.

   Solutions: (1) $P(y = 1)$, (2) $P(x_1 = 1|y = i)$ $(i = 1, 0)$, (3) $P(x_2 = 1|y = i)$ $(i = 1, 0)$, (4) $P(x_3 = 1|y = i)$ $(i = 1, 0)$ (2 pts for correctly naming 1 parameter). 7 independent parameters in total (1 pt for each).

2. **[9 pts.]** What are your estimations for these parameters? (say using standard MLE).

   Solutions: (1) $P(y = 1) = 0.5$,

   (2) $P(x_1 = 1|y = 1) = 0.6$ and $P(x_1 = 1|y = 0) = 0.4$,

   (3) $P(x_2 = 1|y = 1) = 0.2$ and $P(x_2 = 1|y = 0) = 0.4$,

   (4) $P(x_3 = 1|y = 1) = 0.8$ and $P(x_3 = 1|y = 0) = 1$.

   (3 pts for the prior and 1 pt for each of the rest estimation)

3. **[6 pts.]** Now, given a new (test) example $x = (0, 1, 0)$, what is $P(y = 1|x)$? Which class label will the naive Bayes classier assign to this example? Justify your answer.

   Solutions: $P(x|y = 1)P(y = 1) = P(x_1 = 0|y = 1)P(x_2 = 1|y = 1)P(x_3 = 0|y = 1)P(y = 1) = 0.4 \times 0.2 \times 0.2 \times 0.5 = 0.008$. $P(x|y =$

$0)P(y=0) = P(x_1 = 0|y = 0)P(x_2 = 1|y = 0)P(x_3 = 0|y = 0)P(y = 0) = 0.6 \times 0.4 \times 0 \times 0.5 = 0$. Therefore, $P(y = 1|x) = \frac{P(x|y=1)P(y=1)}{P(x|y=1)P(y=1)+P(x|y=0)P(y=0)} = 1$. The assigned label will be $y = 1$. (3pt for the correct $p(y = 1|x)$, 3 pts for the correct label prediction)

4. **[6 pts.]** Now, given a new (test) example $x = (1, 0, 1)$, what is $P(y = 1|x)$? Which class label will the naive Bayes classier assign to this example? Justify your answer.

   Solutions: $P(x|y = 1)P(y = 1) = P(x_1 = 1|y = 1)P(x_2 = 0|y = 1)P(x_3 = 1|y = 1)P(y = 1) = 0.6 \times 0.8 \times 0.8 \times 0.5 = 0.192$. $P(x|y = 0)P(y = 0) = P(x_1 = 1|y = 0)P(x_2 = 0|y = 0)P(x_3 = 1|y = 0)P(y = 0) = 0.4 \times 0.6 \times 1 \times 0.5 = 0.12$. Therefore, $P(y = 1|x) = \frac{P(x|y=1)P(y=1)}{P(x|y=1)P(y=1)+P(x|y=0)P(y=0)} = \frac{0.192}{0.192+0.12} = 8/13 > 0.5$. The assigned label will be $y = 1$. (3pt for the correct $p(y = 1|x)$, 3 pts for the correct label prediction)

5 Gaussian Bayes Classifier and Naive Gaussian Bayes Classifier (10 points)

Consider a binary classification task, where the feature vector X has d dimensions and the class label y is either 1 or 0. We consider two types of classifiers, i.e., (1) Gaussian Bayes Classifier and (2) Gaussian Naive Bayes Classifier. In Gaussian Bayes Classifier, we assume $P(X|y)$ follows a multi-variate Gaussian distribution for each class label ($y = 0$ and $y = 1$). In Gaussian Naive Bayes Classifier, we assume that for each class label ($y = 0$ and $y = 1$), different dimensions of the feature vector X are conditionally independent with each other, each following a single variate Gaussian distribution.

1. **[5 pts.]** How many independent parameters are there in your Gaussian Bayes Classifier?

   Solutions: $(d + (1 + 2 + ... + d)) \times 2 + 1 = d^2 + 3d + 1$ (1 for class prior. for each class label, we have d for mean vector of Gaussian and $(1 + 2 + ... + d)$ for the co-variance matrix. lose two points if treating d^2 parameters for the co-variance matrix. lose 1 point if missing the parameter for the class prior or treating 2 parameters for the class prior.)

2. **[5 pts.]** How many independent parameters are there in your Naive Gaussian Bayes Classifier?

   Solutions: $(1 + 1) \times d \times 2 + 1 = 4d + 1$ (1 for class prior. for each class label, for each dimension of the feature, we have 1 for mean of Gaussian and (1) for the variance. lose 1 point if missing the parameter for the class prior or treating 2 parameters for the class prior.)

Blank Page