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.)

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.

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?

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 $y = 2$?

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

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

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

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

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

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

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 $y = 2$?

- **[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 $y = 2$).

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

- **[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?

- **[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?

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

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

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.

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

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.

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.

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?

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

Blank Page