Sample questions for Test 2 for CSI 431

October 22, 2020

1. **Decision Trees [5pts]** True or False: (a) High entropy means that the partitions in classification are pure. (b) Multiway split of a categorical attribute generally results in more pure partitions than a binary split.

2. **Decision Trees [10pts]** Given the following table:

Point	Age	Car	Risk
x1	25	Sports	L
x2	20	Vintage	H
x3	25	Sports	L
x4	45	SUV	H
x5	20	Sports	H
x6	25	SUV	H

   construct a decision tree using a purity threshold of 100%. Use information gain as the split point evaluation measure. Next, classify the point (Age=27,Car=Vintage).

3. **Decision Trees [10pts]** Given the dataset in the table, answer the following questions:

Instance	a1	a2	a3	Class
1	T	T	5.0	Y
2	T	T	7.0	Y
3	T	F	8.0	N
4	F	F	3.0	Y
5	F	T	7.0	N
6	F	T	4.0	N
7	F	F	5.0	N
8	T	F	6.0	Y
9	F	T	1.0	N

   (a)[5pts] Show which decision will be chosen at the root of the decision tree using information gain, Gini index, and CART measures. Show all split points for all attributes. Note, this is a lot of work and you will not be asked to do as much work on a test. Feel free to use some scripting or excel to complete this in case you want to test all splits

(b)[5pts] What happens to the purity if we use Instance as another attribute? Do you think this attribute should be used for a decision in the tree?

4. **LDA [20pts]**. Consider the data shown in the table:
Answer the following questions:

i	x_i	y_i
x_1	$(4, 2.9)$	1
x_2	$(3.5, 4)$	1
x_3	$(2.5, 1)$	-1
x_4	$(2, 2.1)$	-1

(a)[5pts] Compute μ_{+1} and μ_{-1}, and B, the between-class scatter matrix.
(b)[5pts] Compute S_{+1} and S_{-1}, and S, the within-class scatter matrix.
(c)[5pts] Find the best direction w that discriminates between the classes.
Use the fact that the inverse of $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is given as $A^{-1} = \frac{1}{det(A)} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.
(d)[5pts] Having found the direction w, find the point on w that best separates the two classes.

5. **Bayes v.s. Naive Bayes [25pts]** Consider two classes with the same priors $P(c_1) = P(c_2)$ the same covariance matrix: $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and means $\mu_1 = (1, 1)^T$ and $\mu_1 = (2, 1)^T$. Consider also the following test points: $x = (0, 1)^T$ and $y = (1.5, 0)^T$.

   (a) **[20 pts.]** How would x and y be classified according to full Bayes classifier? Show your computation. Note that the pdf of multivariate normal distributions is $P(x|c_i)\tilde{N}(\mu_i, \Sigma_i)$ with PDF:

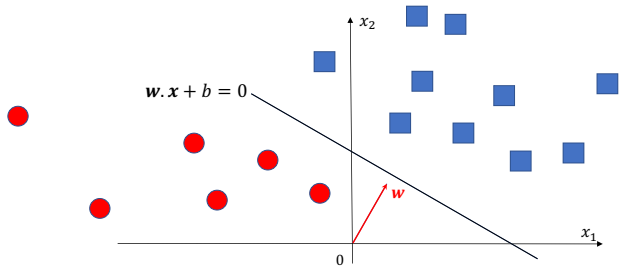
   $$f_i(x) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_i|}} exp\{-\frac{(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)}{2}\}.$$

   Note also that $\frac{1}{\sqrt{(2\pi)^k |\Sigma_i|}}$ will be the same for both classes. The inverse of the covariance $\Sigma^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \Sigma$.

   (b) **[5 pts.]** Would the classification result differ if we use Naive Bayes? Explain why.

6. **Bayesian [15pts]** Q3 after Bayesian (on page 482).
Extensions:
(1)[+5pts] Using the same parameters, classify the same point using Naive Bayes.
(2)[+10pts] Come up with a few samples that would result in the estimated statistical properties of class 1.

7. **SVM [15pts]** Assume that the hyperplane shown in the figure below is canonical w.r.t. the data (samples of two classes squares and circles).



   (a) (5pts) Draw the support hyper planes and and show their equations.

   (b) (5pts) How are the support vectors defined. Show them on the figure.

   (c) (5pts) What is the positive and what is the negative class?

8. **Distance from a point to a hyperplane [10pts]** Consider the plane $h(x)$ defined by $w = (1, -1, 3)$ and $b = 10$.

   (a) (3pts) What is the distance from point $x = (2, 2, 2)$ to h(x)? Explain your computation?

   (b) (3pts) Is $x = (2, 2, 2)$ in the positive or negative half-space defined by $h(x)$?

   (c) (4pts) Give examples of points that could be support vectors for h(x).

9. **Best hyperplane for 3 points [5pts]** Assume there are only 3 point in the training set:

$$\begin{bmatrix} x_1 & x_2 & y \\ 1 & 3 & +1 \\ 3 & 0 & +1 \\ 4 & 4 & -1 \end{bmatrix}$$

   What will be the equation of the optimal canonical hyperplane found by hard margin SVM? What would be the equations of the support hyperplanes? Explain your derivation.

10. **Better margin [5pts]** Which of the following two hyperplanes have a better margin? Compute the margins and explain your reasoning.

$$h_1(x) = x_1 - 2x_2 + 5x_3 - 8 = 0$$
$$h_2(x) = 2x_1 - x_2 + x_3 + 22 = 0$$

11. **Classify using SVM [5pts]** How would the following test points be classified by the SVM model: $w = (-2, 1)$ $b = 3$:
$$\begin{bmatrix} 1 & 3 \\ 3 & 0 \\ 4 & 4 \end{bmatrix}$$

12. **Increase the margin [5pts]** What would be the SVMs with the same separating hyperplane as $w = (-2, 1)$ $b = 3$ but with twice bigger and twice smaller margins?

13. **Soft margin [5pts]** Let $w = (-2, 1)$ $b = 3$ be the separating hyperplane of an SVMs with soft margin. What would be the slack variables for the following training points:

$$\begin{bmatrix} 1 & 3 & +1 \\ 3 & 0 & +1 \\ 4 & 4 & +1 \end{bmatrix}$$

Hint: The slack variables are based on the distances from the support hyperplanes and not the separating hyperplane.

14. **Confusion matrix for classifiers [20pts]** Consider the classifiers on slide 7 (Decision Tree) and slide 11 (SVM) of the Classifier evaluation slides on Blackboard and the corresponding test dataset.

    (a) Compute the corresponding confusion matrices for the classifiers.

    (b) Compute the precision, recall and F-measure of the classifiers.

    (c) What would be the above statistics for the DT classifier if it had only the first split?

15. **Confusion matrix examples [10pts]** Is it possible that the precision of the positive class be 5 times higher than the recall? How about 5 times lower? Give example confusion matrices for the above or argue that they are impossible.

16. **ROC and Precision-recall [30pts]** Consider the SVM classifier from slide 11 of the Classifier evaluation slides and the corresponding testing set.

    (a) (5pts)Compute the signed distances for each testing point from the separating hyperplane.

    (b) (5pts)Use the above signed distances as the score for classifying class +1 (We defined and discussed this score in the context of ROC). What would be the consecutive threshold values to compute an ROC?

(c) (10pts)Draw the ROC (provide the coordinates of each point on the ROC) and compute the AUC.

(d) (10pts)Use the scores from (b) and the thresholds to to draw a precision-recall curve where on the x axis you have the Recall for class +1 and on the y axis you have the Precision for class +1. Similarly provide the coordinates for all points on this curve.

17. **Generic question:** Pick a slide where we worked out an example in class (only topics for the test), change the example, solve the new example. To post a solution post it as Question11-SlideXXX-TopicYYY, where XXX is the slide number and YYY is the topic (slide deck). Post both the new example and your solution.