

Name:

Practice Midterm Exam

CS 165B – Machine Learning

May 2, 2016

9:30-10:45am

This is a **closed-book** test, with one exception: you may use one 8.5"x11" piece of paper with any notes you wish to write on it (front and back).

Be sure to read each question carefully and provide all the information requested. **If the question asks you to explain, do so!**

Show your work. Write your answers in the spaces provided and, if necessary, on the back of the page. If you use the back, draw an arrow or write "SEE BACK" to make sure the graders don't miss it. If you need more space, attach extra sheets of paper (available at the front).

Exams must be turned in by 10:45am sharp.

Good luck!

Manhattan (L1) distance: $d(x, y) = \sum_{i=1}^d |x_i - y_i|$

Euclidian (L2) distance: $d(x, y) = \|x - y\| = \left(\sum_{i=1}^d (x_i - y_i)^2 \right)^{1/2}$

Minkowski (Lp) distance: $d(x, y) = \left(\sum_{i=1}^d |x_i - y_i|^p \right)^{1/p}$

$$\text{Laplace correction} = \frac{N_i + 1}{|S| + k} \quad \text{m-estimate} = \frac{N_i + m\pi_i}{|S| + m}$$

Algorithm $\text{GrowTree}(D, F)$ – grow a feature tree from training data.

Input : data D ; set of features F .

Output : feature tree T with labelled leaves.

if $\text{Homogeneous}(D)$ **then return** $\text{Label}(D)$;

$S \leftarrow \text{BestSplit}(D, F)$; // e.g., BestSplit-Class (Algorithm 5.2)

split D into subsets D_i according to the literals in S ;

for each i **do**

if $D_i \neq \emptyset$ **then** $T_i \leftarrow \text{GrowTree}(D_i, F)$;

else T_i is a leaf labelled with $\text{Label}(D)$;

end

return a tree whose root is labelled with S and whose children are T_i

Minority class

$$\text{Imp}(\dot{p}) = \min(\dot{p}, 1 - \dot{p})$$

Gini index

$$\text{Imp}(\dot{p}) = 2\dot{p}(1 - \dot{p})$$

Entropy

$$\text{Imp}(\dot{p}) = -\dot{p} \log_2(\dot{p}) - (1 - \dot{p}) \log_2(1 - \dot{p})$$

$\sqrt{\text{Gini index}}$

$$\text{Imp}(\dot{p}) = \sqrt{2\dot{p}(1 - \dot{p})}$$

Total impurity:

$$\text{Imp}(\{D_1, \dots, D_l\}) = \sum_{i=1}^l \frac{|D_i|}{|D|} \text{Imp}(D_i)$$

Multivariate least-squares regression
(homogeneous representation)

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \epsilon$$

$$\hat{\mathbf{w}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

$$= \mathbf{S}^{-1} \mathbf{X}^T \mathbf{y}$$

PAC learning outputs, with probability at least $1 - \delta$, a hypothesis h such that $\text{err}_D < \epsilon$

Bayes Rule:

$$P(H_i | D) = \frac{P(D | H_i) P(H_i)}{P(D)}$$

$$\text{False positive rate (FPR)} = \frac{FP}{N} = \alpha$$

$$\text{Accuracy} = \frac{TP + TN}{P + N} = \left(\frac{P}{P + N} \right) TPR + \left(\frac{N}{P + N} \right) TNR$$

$$\text{False negative rate (FNR)} = \frac{FN}{P} = \beta$$

$$\text{Error rate} = \frac{FP + FN}{P + N}$$

$$\text{True positive rate (TPR)} = \frac{TP}{P}$$

$$\text{Precision} = \frac{TP}{\hat{P}}$$

$$\text{True negative rate (TNR)} = \frac{TN}{N}$$

$$\text{Ranking classifier error rate: } rank\text{-err} = err / PN$$

$$\text{Ranking classifier accuracy: } rank\text{-acc} = 1 - rank\text{-err}$$

Sample Midterm Questions

1. [4 points] What is machine learning? What are the three key components of a learning problem?
2. [2 points] Give an example of an unsupervised discrete learning problem.
3. [4 points] What are the three data sets that are typically used in developing a machine learning algorithm or application, and how is each used?
4. [4 points] Describe inductive learning and deductive learning. Which do we focus on in machine learning and why?
5. [4 points] Briefly discuss the relationship between overfitting and generalization in machine learning. What is likely to have lower error on the training data, a linear model or a higher-order polynomial model? How about on the testing data?
6. [3 points] What is the intrinsic dimensionality of a set of data?
7. [6 points] A test for a new, deadly strain of anthrax (that has no symptoms) is known to be 99.9% accurate. The chances of any random person having this strain are one in

a million. You get tested for anthrax during a routine medical exam, and your test comes back positive. If A is the variable that describes whether you have anthrax (true) or not (false), and T is the variable that describes the output of your anthrax test (true if you test positive, false if you test negative), what is the relatively likelihood that you have anthrax? Use Bayes' Rule.

8. [3 points] Give an example of a predictive machine learning task and an example of a descriptive machine learning task.

9. [4 points] In the basic binary linear classifier (with the linear discriminant function midway between the class centroids), the centroid of our positive class is at $(2, 1, 4)$ and the centroid of our negative class is at $(4, 4, 6)$. The decision boundary is a plane defined by the vector \mathbf{w} and the threshold t . If we now add a new positive sample at $(0, -2, 2)$ to the training data and recompute, how will this affect the placement of the decision plane?

10. [3 points] You are given the probability tables for $P(\text{data} \mid \text{hypotheses})$ and $P(\text{hypotheses})$. You need to choose the best model (hypothesis) from this data. What kind of decision rule is this?

11. [4 points] Why is feature selection often performed in a machine learning problem before learning the model?
12. [3 points] In testing the binary classification model you learned from the training data, you get 70 out of 100 instances correct. 40 of those correctly estimated the positive class (the concept), and the rest correctly estimated the negative class. There were a total of 60 positive examples in the test set. What is the false positive rate? The false negative rate? The precision?
13. [4 points] In learning a classifier, you use a loss function in weighing the effects of various training data instances. If the classification margin of an instance is very high, what should the loss function for that instance be (qualitatively, not a specific value)? If the margin of an instance is very low (negative), what should the loss function for that instance be? If an instance is badly misclassified, what should the loss function for that instance be?
14. You need to estimate prior probabilities for your 5-class classifier, and you have { 10, 8, 14, 9, and 5 } samples from your classes. What are your estimated prior probabilities, using Laplace correction?
15. Describe how a loss function (for classification or regression) can be made robust to outliers.

16. In concept learning, what is the difference between a *hypothesis* and a *concept*?
17. Give an example of a hypothesis that is not learnable by the conjunctive hypothesis space representation.
18. In our conjunctive hypothesis space learning, we generally seek the least general generalization. Compared with a more general generalization, what effect does choosing a less general hypothesis have on our false positive rate?
19. What CHS hypothesis will guarantee a *consistent* concept for any problem?
20. In a decision tree learning routine, a particular node has 8 examples of the positive class and 2 examples of the negative class. What is the impurity measure of that node?

21. A decision tree for concept c has five leaves with the following training examples in each leaf:

L1: (5 pos, 2 neg)

L2: (6 pos, 1 neg)

L3: (3 pos, 4 neg)

L4: (0 pos, 3 neg)

L5: (1 pos, 2 neg)

Using Laplace correction, give the ranking order of the leaves (from highest ranked to lowest).

22. In a univariate linear regression problem, what is the geometric interpretation of the regression coefficient?