ECSE-6610: PR Midterm Exam

March 9th, 2018

Name: ________________ RIN: ________________ Score: ________________

- This is a OPEN BOOK & OPEN NOTE exam. But you cannot access the Internet or use your laptop computer. Do the exam independently.
- There are a total of 100 points in the exam. Plan your work accordingly.
- Write out the steps for all problems to receive the full credit. Use additional pages if necessary.

Problem 1: True or False (20 points)

(a) Given a binary classification scenario with Gaussian class conditionals and equal prior probabilities, the optimal decision boundary will be linear.

○ True ○ False

(b) Changing the priors changes the ROC curve.

○ True ○ False

(c) With k-fold cross-validation, larger k is always better.

○ True ○ False

(d) Cross validation will guarantee that our model does not overfit.

○ True ○ False

(e) A classifier trained on less training data is less likely to overfit.

○ True ○ False

(f) As the number of data points approaches ∞, the error rate of a 1-NN classifier approaches 0.

○ True ○ False

(g) The hyperparameters in a classification model can be η (learning rate) and λ (regularization term).

○ True ○ False

(h) In the primal version of SVM, we are minimizing the Lagrangian with respect to $\mathbf{w}$ and in the dual version, we are minimizing the Lagrangian with respect to α.

○ True ○ False

(i) In SVMs, the values of α_i for non-support vectors are 0.
○ True ○ False

(j) For the dual version of soft margin SVM, the α_i's for support vectors satisfy $\alpha_i > C$.
○ True ○ False

Problem 2: Bayes Decision Theory (10 points)

Suppose we have a two-class problem $(A, \bar{A})$, with a single binary values $(x, \bar{x})$. Assume the prior probability $P(A) = 0.33$. Given the distribution of the samples as shown in the following table, use Bayes Rules to compute the values of posterior probabilities of classes.

Table 1: Data used for Problem 2.

	A	$\bar{A}$
x	248	167
$\bar{x}$	82	503

Problem 3: Parameter Estimation (20 points)

Let samples be drawn by successive, independent selections of a state of nature w_i with unknown probability ρ. Let $z_{ik} = 1$ if the state of nature for the $k-$th sample is w_i and $z_{ik} = 0$ otherwise.

(a) **[7 points]** Show that

$$P(z_{i1}, \dots, z_{in}|\rho) = \prod_{k=1}^{n} \rho^{z_{ik}} (1-\rho)^{1-z_{ik}} \tag{1}$$

(b) **[10 points]** Given the equation above, show that the maximum likelihood estimation for ρ.

$$\hat{\rho} = \frac{1}{n} \sum_{k=1}^{n} z_{ik} \tag{2}$$

(c) **[3 points]** Interpret the meaning of your result in words.

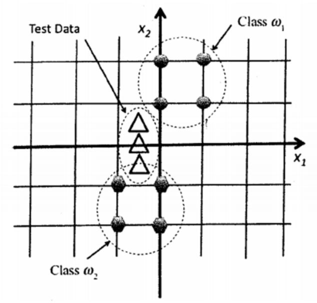
Problem 4 (30 points) In the data below, two sets of objects, truck (ω_1) and cars (ω_2), are observed, In each case, two features are measured $x = [x_1, x_2]$, where x_1 represents the width and x_2 the length.

A training set of four points for each class is observed as follows:

Training set for ω_1: [0, 1], [1, 1], [1, 2], [0, 2]

Training set for ω_2: [0, -1], [-1, -1], [-1, -2], [0, -2] as shown in the figure below.

Figure 1: Data points used for both training and testing.



(a) **[7 points]** Assume Gaussian model for each class, obtain the MLE of the mean μ_i and covariance matrix Σ_i for each class based on the training data, where $i = 1, 2$.

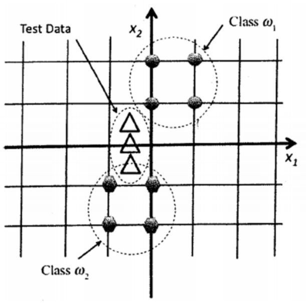
(b) **[5 points]** Using the estimated parameters in (a) and classify the three testing points, based on the likelihood ratio.

(c) **[8 points]** Using the given training data and the Fisher criterion, find the optimal LDA parameters $\mathbf{w}$, and then classify the three testing points based on $\mathbf{w}$.

(d) **[5 points]** Classify the 3 data points based on 1-NN and give their class labels.

(e) **[5 points]** Assume a linear binary SVM is to be constructed from the training data in the figure to classify the two classes, draw the most likely decision boundary, the margin lines and circle the support vectors in the figure below.

Figure 2: Data points used for both training and testing.



Problem 5: Kernel (20 points)

(a) [**10 points**] The polynomial kernel is defined to be

$$k(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + c)^d \tag{3}$$

where $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$, and $c \geq 0$. When we take $d = 2$, this kernel is called the quadratic kernel. Find the feature mapping $\phi(.)$ that corresponds to the quadratic kernel.

(b) [**10 points**] Let k_1 and k_2 be (valid) kernels, *i.e.*, $k_1(\mathbf{x}, \mathbf{y}) = \phi_1(\mathbf{x})^T \phi_1(\mathbf{y})$ and $k_2(\mathbf{x}, \mathbf{y}) = \phi_2(\mathbf{x})^T \phi_2(\mathbf{y})$. Show that $k = k_1 + k_2$ is a valid kernel by explicitly constructing a corresponding feature mapping $\phi(.)$.