GEORGIA INSTITUTE OF TECHNOLOGY
School of Electrical and Computer Engineering

ECE 6254
Midterm Sample Problems

Problem 1 (0 pts):

Short Answer Questions: Answer each problem (in words) using 1-2 short sentences.)

a. Describe the difference between Hoeffding's inequality (as we have stated and used it in this course) and the VC generalization bound.

b. What is the "kernel trick" and why do we use it in SVMs?

c. In training an SVM how do new vectors outside the margin affect the optimization?

d. Explain the difference between classification and regression.

e. The Bayes' classifier is considered the best classifier—in what sense is it the best (i.e., what does it optimize)?

f. If the Bayes' classifier is the best performing classifier, why are we concerned with other classifiers such as the SVM?

g. For a given set of training samples and parameters which of the following yield unique solutions: *Perceptron Learning Algorithm* or *Support Vector Machine*?

h. Why is the observed risk on a training set less than or equal to the risk of the best classifier when measured on that training set?

i. Can a k-nearest neighbor classifier approach the accuracy of a Bayesian classifier? Explain.

Problem 2 (0 pts):

We have the following data points belonging to two classes:

$$\begin{aligned} \omega_1 &= \{(1,1)^T, (1,-2)^T\} \\ \omega_2 &= \{(0,0)^T, (2,0)^T\} \end{aligned}$$

a. Are these classes linearly separable?

b. Let

$$\underline{f} = \begin{bmatrix} x_1^2 \\ x_2^2 \end{bmatrix}$$

In this transformed domain, are the classes linearly separable?

$$\begin{aligned} \omega_1 &= \{(1,1)^T, (1,-2)^T\} \\ \omega_2 &= \{(0,0)^T, (2,0)^T\} \end{aligned}$$

c. Apply the perceptron algorithm to this **transformed** data. Start with $\mathbf{w} = (0,0)^T$ and $b = 0$.

d. Draw the decision surface and the linear discriminant vector, $\mathbf{w}$ found in part **c**.

e. Draw the decision surface in the pre-transformed space.

Problem 3 (0 pts):

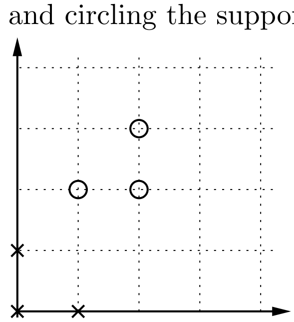
SVMs

We have the following data points belonging to two classes:

$$\begin{aligned} \omega_1 &= \{(1,2)^T, (2,2)^T, (2,3)^T\} \\ \omega_2 &= \{(0,0)^T, (1,0)^T, (0,1)^T\} \end{aligned}$$

Let the class indicator for ω_1 be $t_i = 1$ and the class indicator for ω_2 be $t_i = -1$.

a. Give the SVM solution that separates ω_1 and ω_2 by carefully sketching the separating line and circling the support vectors.



b. Considering **only the first point of each class**,[1] find the Lagrange multipliers, a_1 and a_2 that satisfy the dual problem.

[1] *i.e.*, assume our data classes only have one data sample each.

c. Find $\mathbf{w}$ and b that define the linear separating line based on the Lagrange multipliers in the previous part. Show your work for any partial credit.

Problem 4 (0 pts):

Consider a binary classification problem involving a single (scalar) feature x and suppose that $X|Y=0$ and $X|Y=1$ are continuous random variables with densities given by

$$f_{X|Y}(x|0) = \frac{1}{\sqrt{2}} e^{-\sqrt{2}|x|} \qquad \text{and} \qquad f_{X|Y}(x|1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

respectively.

1. Suppose that $\pi_0 = \mathrm{P}(Y=0) = \frac{1}{2}$ and hence $\pi_1 = \mathrm{P}(Y=1) = \frac{1}{2}$. Derive the optimal classification rule in terms of minimizing the probability of error. [Hint: Drawing a sketch of the two distributions may help you to see what the rule should look like.] Simplify your decision rule as much as possible, but you may leave your answer in terms of quantities like $\sqrt{2}$ and π – you do not need to provide a decimal answer.

2. Give a formula for the Bayes risk for this problem. Your answer may be stated in terms of integrals of the distributions $f_{X|Y}(x|0)$ and $f_{X|Y}(x|1)$, i.e., you do not have to evaluate the integrals (but you do need to give me the endpoints of each integral).

Problem 5 (0 pts):

In class we showed that the solution to the least-squares problem

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \ \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2$$

is given by $\widehat{\boldsymbol{\theta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$. Note that in the special case where $\mathbf{A} = \mathbf{I}$, this reduces to the natural estimate of $\widehat{\boldsymbol{\theta}} = \mathbf{y}$.

In Tikhonov regularization we consider the alternative optimization problem of

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \ \|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|_2^2 + \lambda\|\boldsymbol{\theta}\|_2^2.$$

Give a formula for the solution to this problem in the special case of $\mathbf{A} = \mathbf{I}$. Simplify your answer as much as possible.

Problem 6 (0 pts):

Consider the set of classifiers defined by *positive rings*, i.e., classifiers who can be defined for any $\mathbf{x} \in \mathbb{R}^2$ via,

$$h(\mathbf{x}) = \begin{cases} +1 & \text{if } r_1 \leq \|\mathbf{x} - \mathbf{c}\|_2 \leq r_2 \\ -1 & \text{otherwise,} \end{cases}$$

for some center $\mathbf{c} \in \mathbb{R}^2$ and radii $r_1, r_2 \geq 0$.

a. Show that the VC dimension d_{VC} of the set of all such classifiers is *at least* 4.

b. Describe in a few sentences what you would need to do in a proof that $d_{\text{VC}} < 5$, which together with part (a) would prove that in fact $d_{\text{VC}} = 4$. Do not simply restate the definition of the VC dimension, but describe what you would need to do in the proof.

Problem 7 (0 pts):

Suppose we know that a set of classifiers $\mathcal{H}$ shatters a *particular* set of points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ in $\mathbb{R}^d$. What do we know about the VC dimension of $\mathcal{H}$? Circle one:

- $d_{\text{VC}}(\mathcal{H}) = m$
- $d_{\text{VC}}(\mathcal{H}) \geq m$
- $d_{\text{VC}}(\mathcal{H}) \leq m$
- none of the above