CSE 569 Homework #4
Total 3 points
Due Tuesday, Nov. 17 **by 11:59pm**.

Question 1. **(Problem 1 of Chapter 5)**

Explore the applicability of linear discriminants for unimodal and multimodal problems in two dimensions through the following.

(a) Sketch two multimodal distributions for which a linear discriminant could give excellent or possibly even the optimal classification accurcy.

(b) Sketch two unimodal distributions for which even the best linear discriminant would give poor classification accuracy.

(c) Consider two circular Gaussian distributions $p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, a_i\mathbf{I})$ and $P(\omega_i)$ for $i = 1, 2$ where $\mathbf{I}$ is the identity matrix and the other parameters can take on arbitrary values. Without performing any explicit calculations, explain whether the optimal decision boundary for this two-category problem must be a line. If not, sketch an example where the optimal discriminant is not a line.

Question 2. **(Problem 2 of Chapter 5)**.

Consider a linear machine with discriminant functions $g_i(\mathbf{x}) = \mathbf{w}^t\mathbf{x} + w_{i0}$, $i = 1, \ldots, c$. Show that the decision regions are convex by showing that if $\mathbf{x}_1 \in \mathcal{R}_i$ and $\mathbf{x}_2 \in \mathcal{R}_i$ then $\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2 \in \mathcal{R}_i$ for $0 \leq \lambda \leq 1$.

Question 3. **(Problem 9 of Chapter 5)**.

The *convex hull* of a set of vectors $\mathbf{x}_i$, $i = 1, \ldots, n$ is the set of all vectors of the form

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i,$$

where the coefficients α_i are nonnegative and sum to one. Given two sets of vectors, show that either they are linearly separable or their convex hulls intersect. (To answer this, suppose that both statements are true, and consider the classification of a point in the intersection of the convex hulls.)

Question 4. Consider 2-class, linear classification, with n input samples $\{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$. In the Perceptron training rule, we update the weight vector **a** by an amount proportional to the sum of the (normalized) misclassified patterns. A single-sample version will also work, which updates **a** by adding to it any randomly chosen misclassified pattern, resulting the following algorithm in pseudo-code,

Algorithm-Single: Fixed-increment single-sample Perceptron

```
initialize  a, k = 0
do k = (k+1) modulo n
        if  y_k is misclassified by a  then a = a + y_k
until all patterns properly classified
return a
```

Now, let's consider a new rule to update the weight vector using only the current *worst*-classified pattern(s) from each class. Given a current weight vector, the *worst*-classified pattern for a class may be *either* (a) "the one on the wrong side of the decision plane *and* furthest to the plane", *or*, if all the patterns are already on the correct side, (b) "the one on the correct side of the decision plane *and* closest to the plane (including the special case 'right on the plane, i.e. distance-to-the-plane =0')".

Write an algorithm in terms of pseudo-code to implement the above new rule. Your algorithm should be similar to the above Algorithm-Single in terms of level of detail, with the following special requirements: (i) You need to explicitly show how the *worst*-classified pattern is determined in each iteration; and (ii) You need to specify when the algorithm will terminate.

Question 5. Given the following training samples from two classes:

Class 1: $(1, 1)^t, (2, 2)^t, (2, 0)^t$

Class 2: $(0, 0)^t, (1, 0)^t, (0, 1)^t$

(1) Are the two classes linearly separable?

(2) Use the Pseudoinverse Approach to find two linear decision boundaries by using the following two vectors for **b**.

$(1, 1, 1, 1, 1, 1)^t, \quad (1, 1, 1, 1, 1, 2)^t,$

(You may use any tool to computer matrix inversion.)

(3) Verify if the solutions from (2) can indeed classify the samples correctly.

Question 6. Given the same training set as in Problem 5.

(1) Plot the samples in the feature space, and construct by inspection the max margin linear classifier, which would be what you will get by SVM.

(2) What is the margin of the SVM classifier?

(3) Plot the decision line from (1) together with those found from Problem 5.