ELE 535: Machine Learning and Pattern Recognition
Homework 1

Zachary Hervieux-Moore

Monday 24$^{\text{th}}$ September, 2018

Exercise 1: Consider the training data $\{(x_j, y_j)\}_{j=1}^m$, with $x_j \in \mathbb{R}^n$ and $y_j \in \{0, 1\}$. Assume the training data has an equal number of examples from each class. Hence the estimated prior probabilities of each class are equal. The nearest centroid classifer has

$$\hat{y}(x) = \begin{cases} 1, & \text{if } \|x - \hat{\mu}_1\|_2 < \|x - \hat{\mu}_0\|_2; \\ 0, & \text{o.w.} \end{cases}$$

a) Show that the nearest centroid classifier is a linear classifier with

$$w = (\hat{\mu}_0 - \hat{\mu}_1)$$
$$b = (\hat{\mu}_1 - \hat{\mu}_0)^T \frac{(\hat{\mu}_1 + \hat{\mu}_0)}{2}$$

b) Show that the classifier can also be written as

$$\hat{y}(x) = \begin{cases} 1, & \text{if } \langle \hat{\mu}_0 - \hat{\mu}_1, x - \hat{\mu} \rangle < 0; \\ 0, & \text{o.w.} \end{cases}$$

Here $\hat{\mu} \triangleq \frac{1}{m} \sum_{i=1}^m x_i$ denotes the mean of the training examples. So the result of classification depends solely on the sign of a inner product.

c) By neatly sketching the vectors $x - \hat{\mu}$ and $\hat{\mu}_0 - \hat{\mu}_1$, give a geometric interpretation of this classifier.

d) Suppose we first center the training data by subtracting $\hat{\mu}$ from each training example. Determine the form of the nearest centroid classifier for the centered training data.

Answer:

a) We define a linear classifier as:

$$\begin{aligned} & w^T x + b < 0 \\ \iff & (\hat{\mu}_0 - \hat{\mu}_1)^T x + (\hat{\mu}_1 - \hat{\mu}_0)^T \frac{(\hat{\mu}_1 + \hat{\mu}_0)}{2} < 0 \\ \iff & x^T x - 2\hat{\mu}_1^T x + \hat{\mu}_1^T \hat{\mu}_1 < x^T x - 2\hat{\mu}_0^T x + \hat{\mu}_0^T \hat{\mu}_0 \\ \iff & \|x - \hat{\mu}_1\|_2^2 < \|x - \hat{\mu}_0\|_2^2 \\ \iff & \|x - \hat{\mu}_1\|_2 < \|x - \hat{\mu}_0\|_2 \end{aligned}$$

where I added $x^T x$ to both sides in the second step.

b) By assumption of equal class sizes, we have that

$$\hat{\mu} = \frac{\hat{\mu}_1 + \hat{\mu}_0}{2}$$

Then, our linear classifier from above becomes

$$\begin{aligned}
&(\hat{\mu}_0 - \hat{\mu}_1)^T x + (\hat{\mu}_1 - \hat{\mu}_0)^T \frac{(\hat{\mu}_1 + \hat{\mu}_0)}{2} < 0 \\
\iff& (\hat{\mu}_0 - \hat{\mu}_1)^T x + (\hat{\mu}_1 - \hat{\mu}_0)^T \hat{\mu} < 0 \\
\iff& \langle \hat{\mu}_0 - \hat{\mu}_1, x - \hat{\mu} \rangle < 0
\end{aligned}$$

Which, by equivalence of part a), is equivalent to the nearest centroid classifier.

c) Sketch shown below in Figure 1. Essentially, the geometric interpretation is that if the angle between $x - \hat{\mu}$ and $\hat{\mu}_0 - \hat{\mu}_1$ is obtuse (or the dot product is negative), then the point x lies on the side of the mean closer to $\hat{\mu}_1$ and hence closer to $\hat{\mu}_1$. In my example, the angle is acute and therefore x is closer to $\hat{\mu}_0$ and should be classified as such.

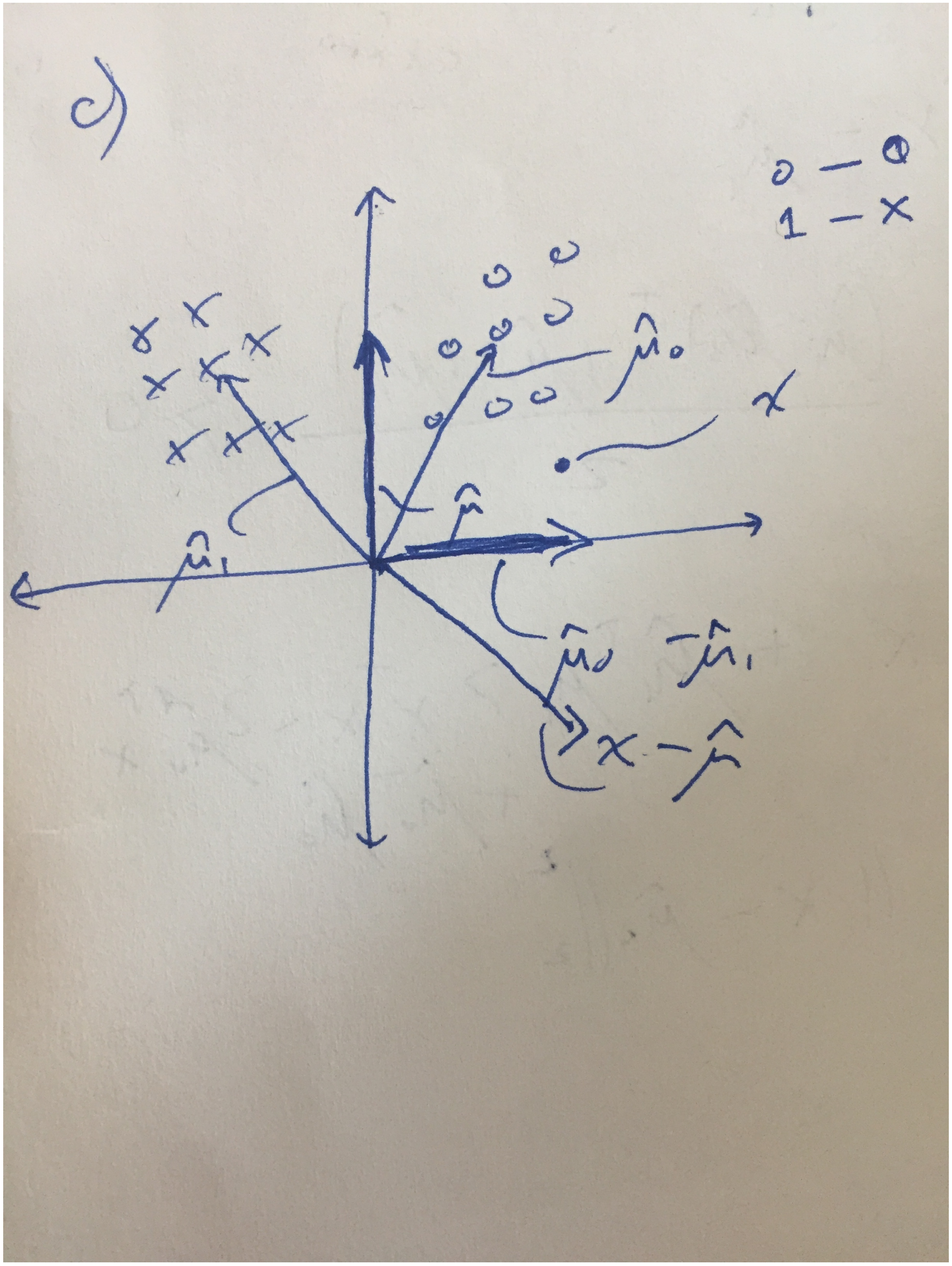

Figure 1: Sktech of $x - \hat{\mu}$ and $\hat{\mu}_0 - \hat{\mu}_1$

d) We define

$$\tilde{x} = x - \hat{\mu}$$

Then, using our characterization from part b), we get

$$\begin{aligned}
&\langle \hat{\mu}_0 - \hat{\mu}_1, x - \hat{\mu} \rangle \\
&= \langle (\hat{\mu}_0 - \hat{\mu}) - (\hat{\mu}_1 - \hat{\mu}), x - \hat{\mu} \rangle \\
&= \langle \tilde{\mu}_0 - \tilde{\mu}_1, \tilde{x} \rangle
\end{aligned}$$

So our nearest centroid classifier becomes

$$\hat{y}(x) = \begin{cases} 1, & \langle \tilde{\mu}_0 - \tilde{\mu}_1, \tilde{x} \rangle; \\ 0, & \text{o.w.} \end{cases}$$

Exercise 2: For a $n \times m$ real matrix X show that:

a) $\mathcal{R}(X) \triangleq \{z\mathbb{R}^n : z = Xw, \text{ for } w \in \mathbb{R}^m\}$ is a subspace of $\mathbb{R}^n$.

b) $\mathcal{N}(X) \triangleq \{a \in \mathbb{R}^m : Xa = \mathbf{0}\}$ is a subspace of $\mathbb{R}^m$.

Answer:

Note: To show S is a subspace, one must show 3 things:

i) $\mathbf{0} \in S$

ii) if $x, y \in S$ then $x + y \in S$

iii) if $x \in S$ and $\alpha \in \mathbb{R}$ then $\alpha x \in S$

a) i) By picking $w = \mathbf{0}_m$ we have that $z = Xw = X\mathbf{0} = \mathbf{0}_n$ and so $\mathbf{0}_n \in \mathcal{R}(X)$

ii) Pick $z_1, z_2 \in \mathcal{R}(X)$ then we have

$$z_1 + z_2 = Xw_1 + Xw_2 = X(w_1 + w_2)$$

and since $(w_1 + w_2) \in \mathbb{R}^m$ we have $z_1 + z_2 \in \mathcal{R}(X)$

iii) Pick $z \in \mathcal{R}(X)$ and $\alpha \in \mathbb{R}$ then we have

$$\alpha z = \alpha Xw = X(\alpha w)$$

and since $(\alpha w) \in \mathbb{R}^m$ we have $\alpha z \in \mathcal{R}(X)$

b) i) By picking $a = \mathbf{0}_m$ we have that $Xa = X\mathbf{0}_m = \mathbf{0}_n$ and so $\mathbf{0}_m \in \mathcal{N}(X)$

ii) Pick $a_1, a_2 \in \mathcal{N}(X)$ then we have

$$X(a_1 + a_2) = Xa_1 + Xa_2 = \mathbf{0}_n + \mathbf{0}_n = \mathbf{0}_n$$

and so we have $a_1 + a_2 \in \mathcal{N}(X)$

iii) Pick $a \in \mathcal{N}(X)$ and $\alpha \in \mathbb{R}$ then we have

$$X(\alpha a) = \alpha Xa = \alpha \mathbf{0}_n = \mathbf{0}_n$$

and so we have $\alpha a \in \mathcal{N}(X)$

Exercise 3:

a) Let $A_j \in \mathbb{R}^{n_j \times m}$ and $\mathcal{N}_j = \{x \in \mathbb{R}^m : A_j x = \mathbf{0}\}$, $j = 1, 2$. Show that $\mathcal{N}_1 \cap \mathcal{N}_2$ is a subspace of $\mathbb{R}^m$. Give a similar matrix equation for this subspace.

b) Let $A_j \in \mathbb{R}^{n \times m_j}$ and $\mathcal{R}_j = \{y\mathbb{R}^n : y = A_j x, \text{ with } x \in \mathbb{R}^{m_j}\}$, $j = 1, 2$. Show that $\mathcal{R}_1 + \mathcal{R}_2 = \{y_1 + y_2 : y \in \mathcal{R}_1, y_2 \in \mathcal{R}_2\}$ is a subspace of $\mathbb{R}^n$. Give a similar matrix equation for this subspace.

Answer:

Note: To show S is a subspace, one must show 3 things:

i) $\mathbf{0} \in S$

ii) if $x, y \in S$ then $x + y \in S$

iii) if $x \in S$ and $\alpha \in \mathbb{R}$ then $\alpha x \in S$

a) i) By question Q2 part b), we know that $\mathbf{0}_m \in \mathcal{N}_j$ for $j = 1, 2$ and so $\mathbf{0}_m \in \mathcal{N}_1 \cap \mathcal{N}_2$.

ii) Pick $x_1, x_2 \in \mathcal{N}_1 \cap \mathcal{N}_2$ then we have

$$A_j(x_1 + x_2) = A_j x_1 + A_j x_2 = \mathbf{0}_n$$

For both $j = 1, 2$ and since $(x_1 + x_2) \in \mathbb{R}^m$ we have $x_1 + x_2 \in \mathcal{N}_1 \cap \mathcal{N}_2$

iii) Pick $x \in \mathcal{N}_1 \cap \mathcal{N}_2$ and $\alpha \in \mathbb{R}$ then we have

$$A_j(\alpha x) = \alpha A_j x = \alpha \mathbf{0}_n = \mathbf{0}_n$$

For both $j = 1, 2$ and since $(\alpha x) \in \mathbb{R}^m$ we have $\alpha x \in \mathcal{N}_1 \cap \mathcal{N}_2$

The similar matrix equation is

$$\mathcal{N}_1 \cap \mathcal{N}_2 = \left\{ x \in \mathbb{R}^m : \begin{bmatrix} A_1 & \mathbf{0}_{n_1 \times m} \\ \mathbf{0}_{n_2 \times m} & A_2 \end{bmatrix} \begin{bmatrix} x \\ x \end{bmatrix} = \mathbf{0}_{n_1 + n_2} \right\}$$

b) i) By question Q2 part a), we know that $\mathbf{0}_n \in \mathcal{R}_j$ for $j = 1, 2$ and so $\mathbf{0}_n \in \mathcal{R}_1 + \mathcal{R}_2$ as $\mathbf{0}_n + \mathbf{0}_n = \mathbf{0}_n$.

ii) Pick $a, b \in \mathcal{R}_1 + \mathcal{R}_2$ then we also have for some $a_1, b_1 \in \mathcal{R}_1$ and $a_2, b_2 \in \mathcal{R}_2$ where

$$\begin{aligned} a &= a_1 + a_2 = A_1 x_1 + A_2 x_2 \\ b &= b_1 + b_2 = A_1 x_1' + A_2 x_2' \\ a + b &= a_1 + b_1 + a_2 + b_2 = A_1(x_1 + x_1') + A_2(x_2 + x_2') \end{aligned}$$

As all the dimensions are appropriate and everything is in the proper spaces, we have $a + b \in \mathcal{R}_1 + \mathcal{R}_2$.

iii) Pick $a \in \mathcal{R}_1 + \mathcal{R}_2$ and $\alpha \in \mathbb{R}$ then for some $a_1 \in \mathcal{R}_1$ and $a_2 \in \mathcal{R}_2$

$$\begin{aligned} a &= a_1 + a_2 = A_1 x_1 + A_2 x_2 \\ \alpha a &= \alpha A_1 x_1 + \alpha A_2 x_2 = A_1(\alpha x_1) + A_2(\alpha x_2) \end{aligned}$$

As all the dimensions are appropriate and everything is in the proper spaces, we have $\alpha a \in \mathcal{R}_1 + \mathcal{R}_2$.

The similar matrix equation is

$$\mathcal{R}_1 + \mathcal{R}_2 = \left\{ y \in \mathbb{R}^n : y = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \text{ for } \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \in \mathbb{R}^{m_1 + m_2} \right\}$$

Exercise 4: For $A, B \in \mathbb{R}^{m\times n}$, the inner product of A and B is $\langle A, B\rangle \triangleq \sum_i \sum_j A_{i,j} B_{i,j}$. Show that $\langle A, B\rangle = \text{trace}(A^T B)$.

Answer: We do this by direct computation.

$$\begin{aligned} \text{trace}(A^T B) &= \sum_j (A^T B)_{j,j} = \sum_j \sum_i A^T_{j,i} B_{i,j} \\ &= \sum_j \sum_i A_{i,j} B_{i,j} = \sum_i \sum_j A_{i,j} B_{i,j} = \langle A, B\rangle \end{aligned}$$

Where the sum can be exchanged because they are both finite.

Exercise 5: Consider the vector space of real $n \times n$ matrices. Let $\mathcal{S}$ and $\mathcal{A}$ denote the subsets of symmetric ($P^T = P$) and antisymmetric ($A^T = -A$) matrices, respectively. Show that $\mathcal{S}$ and $\mathcal{A}$ are subspaces of $\mathbb{R}^{n\times n}$ and that $\mathcal{S}^\perp = \mathcal{A}$. Hence $\mathbb{R}^{n\times n} = \mathcal{S} \oplus \mathcal{A}$

Answer:

Note: To show S is a subspace, one must show 3 things:

i) $\mathbf{0} \in S$

ii) if $x, y \in S$ then $x + y \in S$

iii) if $x \in S$ and $\alpha \in \mathbb{R}$ then $\alpha x \in S$

We start with symmetric matrices.

i) The all 0 matrix is clearly symmetric.

ii) Pick $P_1, P_2 \in \mathcal{S}$ then we have that

$$(P_1 + P_2)_{i,j} = (P_1 + P_2)_{j,i}$$

since $P_{1_{i,j}} = P_{1_{i,j}}$ and $P_{2_{i,j}} = P_{2_{i,j}}$ and so $P_1 + P_2 \in \mathcal{S}$.

iii) Pick $P \in \mathcal{S}$ and $\alpha \in \mathbb{R}$ then we have that

$$(\alpha P)^T = \alpha P^T = \alpha P$$

and so $\alpha P \in \mathcal{S}$

So symmetric matrices form a subspace. Very similarly for antisymmetric matrices.

i) The all 0 matrix is clearly antisymmetric.

ii) Pick $A_1, A_2 \in \mathcal{A}$ then we have that

$$(A_1 + A_2)_{i,j} = -(A_1 + A_2)_{j,i}$$

since $A_{1_{i,j}} = -A_{1_{i,j}}$ and $A_{2_{i,j}} = -A_{2_{i,j}}$ and so $A_1 + A_2 \in \mathcal{S}$.

iii) Pick $A \in \mathcal{A}$ and $\alpha \in \mathbb{R}$ then we have that

$$(\alpha A)^T = \alpha A^T = -\alpha A$$

and so $\alpha A \in \mathcal{A}$.

So antisymmetric matrices form a subspace. Now to show that they are orthogonal complements of each other. We first define the complement of $\mathcal{S}$.

$$\mathcal{S}^{\perp} \triangle \{X : \forall\ P \in \mathcal{S}, \langle X, P \rangle = \mathbf{0}\}$$

From this, it is trivial to see that $\mathcal{A} \subseteq \mathcal{S}^{\perp}$ because the inner product of a symmetric and antisymmetric matrix will yield $\mathbf{0}$ by the definition of the sum shown in question 4. Similarly, showing $\mathcal{S}^{\perp} \subseteq \mathcal{A}$ is easy to see from the summation. For $\langle X, P \rangle = \mathbf{0}$, we need

$$\sum_i \sum_j X_{i,j} P_{i,j} = \mathbf{0}$$

By picking the all 0's symmetric matrix, we get that the diagonal of $X = 0$. Furthermore, by picking symmetric matrices with 0's every where except for one entry (and its pair) to be 1, we get that $X_{i,j} + X_{j,i} = 0$ or $X_{i,j} = -X_{j,i}$. Thus, we have that $X \in \mathcal{A}$. Thus we have shown both sides are subsets of each other and conclude that $\mathcal{A} = \mathcal{S}^{\perp}$ and that this means $\mathbb{R}^{n \times n} = \mathcal{S} \oplus \mathcal{A}$.