ECE 174 Homework # 4 – Due Tuesday 11/21/17

NOTE: *Do this homework set as soon as possible.* Otherwise you will be short of time at the end of the quarter because Homework 5 and the second computer assignment *are both due on the last lecture of the quarter.*

Computer Project 2 Due Date

The second computer project is due the very last lecture of the quarter, **Thursday, December 7, 2017**. You should begin to read the assignment now so that you can follow my explanations in class.

Reading

Meyer §4.6 and §5.13; the *singular value decomposition* (SVD) description on pages 412–414 (ignore example 5.12.1); the SVD examples 5.12.3 and 5.13.4 on pages 418–421; and the discussion of the pseudoinverse and the SVD on pages 422–424.

A summary of material gleaned from the text and sources is given in the lecture supplement on Hilbert space theory and the lecture slides located on the class website.

If you want to understand the determinant function as a "volume operator" (this is optional, but highly recommended if you want to "demystify" the determinant) you should carefully study the material presented in Meyer Example 5.13.2 on pages 431–433 (especially the connection to Gram–Schmidt and QR) and Meyer Example 6.1.4 on pages 468–469. An excellent discussion of determinants as volume operators can be found in the textbook *Linear Algebra: an Introductory Approach* by C.W. Curtis, which can be checked out from the S&E library.

Homework

1. Thinking geometrically (i.e., in terms of the four fundamental subspaces) prove Fredholm's Alternative: Given the system $Ax = b \neq 0$. One, and only one, of the following two statements is true:

   a) $Ax = b$ has a solution.
   b) $\langle b, y \rangle \neq 0$ with $A^* y = 0$.

2. Meyer Problem 4.6.7. Set up an solve using the geometric perspective described in the class and in section 5.13 of Meyer. Interpret the variable x as "time" in units of days (say, before and after a "zero day") and y as a measure quantity to be fit for predictive purposes (say the price per share of some stock you are interested in). Use both of your fits to predict the value of a share of stock on day 6.

3. The Perceptron, Linear Classifiers, & Separating Hyperplanes.

As you will come to learn, the *linear separating hyperplane classifier* described below is *one of the fundamental inference engines*[1] used in pattern classification and detection. In subsequent courses you will learn when this simple classifying is optimal in a Bayesian sense and you will learn that in general situations near-optimal performance can be used in its role as a key component of a support vector machine (SVM).

A "hard decision function" (or "hard limiter" or "hard threshold device") is given by

$$\phi(z) = \begin{cases} +1, & z \geq 0 \\ -1, & z < 0 \end{cases},$$

where $z \in \mathbb{R}$ is any real–valued scalar.

A *perceptron* is a simple *artificial neural network* defined by

$$N(x) = \phi(\omega^T x - d),$$

where $\omega, x \in \mathbb{R}^n$ ($\omega \neq 0$) and $d \in \mathbb{R}$. Note that

$$N(x) = \begin{cases} +1, & \omega^T x - d \geq 0 \\ -1, & \omega^T x - d < 0 \end{cases}.$$

The vector x is interpreted to be a *feature vector*, which is a point in the n–dimensional *feature space* $\mathcal{X} = \mathbb{R}^n$. Given a measured feature vector, x, the two possibilities $N(x) = \pm 1$ indicates which of two mutually exclusive and fully exhaustive events is the case.[2] In this homework problem you will give a feature–space geometric interpretation to the function,

$$h(x) = \omega^T x - d : \mathbb{R}^n \to \mathbb{R}\,,$$

and the condition $h(x) \geq 0$ (corresponding to the "yes" decision $N(x) = +1$) versus the condition $h(x) < 0$ (corresponding to the "no" decision $N(x) = -1$).

An *affine subspace* (or *linear variety* or *linear manifold*) is defined to be a translated subspace with the dimension of the affine subspace defined to be the dimension of the translated subspace.[3] In particular, a *hyperplane* is an $n-1$ dimensional affine subspace in the space $\mathbb{R}^n$. A hyperplane partitions the space into two disjoint regions plus the

[1] And arguably *the* fundamental inference engine.

[2] E.g., "it is the case that the stock market will go up tomorrow" versus "it is the case that the stock market will not go up tomorrow." In this situation, the components of the vector x could be a variety of economic and industrial indicators. The parameters ω and d will have been determined by fitting to prior data.

[3] We are already acquainted with the least-squares solution affine subspace $x_p + \mathcal{N}(A)$, where x_p is any particular least-squares solution (which must satisfy the normal equations) for the system $y = Ax$. The least-squares solution set $x_p + \mathcal{N}(A)$ is the null space of A translated by any particular least-squares solution.

points on the hyperplane itself, and therefore all points in $\mathbb{R}^n$ not on the hyperplane must fall on either side of the hyperplane.

If there are two disjoint subsets, say C_1 and C_2, of the space which each lie on opposite sides of the hyperplane, we refer to the hyperplane as a *separating hyperplane* which separates the two subsets, and we say that C_1 and C_1 are *linearly separable*. Suppose we wish to classify a feature vector x according to whether x belongs to class C_1 or to class C_2. If the two classes are linearly separable, then we can equivalently classify x according to which side of a separating hyperplane x belongs to. This homework problem shows that this decision is given by determining the value of $N(x) = \pm 1$.

(a) Show that the set $\mathcal{H} = \left\{ x \,\middle|\, \omega^T x - d = 0 \right\}$ is an $n-1$ dimensional affine subspace and hence is a hyperplane in $\mathbb{R}^n$.[4]

(b) Find the minimum 2–norm vector x_0 from the origin to the hyperplane $\mathcal{H}$.[5]

(c) Show that the hyperplane can also be written as

$$\mathcal{H} = \{ x \,|\, D(x) = 0 \}$$

where

$$D(x) = n^T x - \Delta \,,$$

with n a unit vector perpendicular to $\mathcal{H}$ and Δ the signed distance along x_0 from the origin to $\mathcal{H}$. Write n and Δ in terms of ω and d and explain why we can write the perceptron equivalently as

$$N(x) = \phi(D(x)) \,.$$

(d) Note that $\mathcal{H}$ and the unit vector n together[6] define a "+"–halfspace, $\mathcal{S}^+$, and a "−"–halfspace, $\mathcal{S}^-$, if we define $\mathcal{S}^+$ to be the halfspace that n points *into* if its base is set on $\mathcal{H}$ and $\mathcal{S}^-$ to be the halfspace that $-n$ points into. Also note that $\mathcal{S}^-$, $\mathcal{H}$ and $\mathcal{S}^+$ are disjoint while $\mathbb{R}^n = \mathcal{S}^- \cup \mathcal{H} \cup \mathcal{S}^+$.[7]

(i) Show that $\mathcal{S}^+ = \{ x \,|\, N(x) = +1 \}$ and $\mathcal{S}^- = \{ x \,|\, N(x) = -1 \}$.

(ii) Show that the function $D(x) = n^T x - \Delta$ can be interpreted as a "distance function" which give the degree of penetration of a point x into either $\mathcal{S}^+$ (if

[4]Note that the solutions to $\omega^T x - d = 0$ are solutions to the full row-rank system $Ax = y$ (taking $A = \omega^T$ and $y = d$) and thus we can also make the identification $\mathcal{H} = x_p + \mathcal{N}(A) = x_p + \mathcal{N}(\omega^T)$ for any particular solution (point on the hyperplane) x_p.

[5]I.e., find the minimum distance from the origin to the hyperplane. Note that this problem can be written as

$$\min_x \|x\|^2 \quad \text{subject to} \quad \omega^T x = d \,.$$

[6]Together $(\mathcal{H}, n)$ is called an *oriented* hyperplane.

[7]That is, $\mathcal{S}^-$, $\mathcal{H}$ and $\mathcal{S}^+$ form a *partition* of the space.

$D(x) > 0$) or $\mathcal{S}^-$ (if $D(x) < 0$) where "degree of penetration" means the *minimum* distance from the point x to the hyperplane $\mathcal{H}$. What information about a feature point x does the perceptron $N(x)$ throw away?

Because our decision is based on projecting the feature x along the line defined by the "direction" n, a procedure which makes a classification based on the value of $D(x)$ is known as a *linear classifier.* The linear classifier (which, as we show above, is equivalent to the use of a separating hyperplane as a decision boundary in feature space) is a fundamental tool used in pattern classification and computer-based decision making.

4. **SVD by hand using geometric reasoning.** Using paper and pencil *only* (do *not* use Matlab or a calculator) for each of the following matrices compute the following:[8] (*i*) The SVD, the rank of the matrix and the dimensions of each of the four fundamental subspaces; (*ii*) Orthonormal bases for each of the four fundamental subspaces; (*iii*) From the SVD, the orthogonal projection operator onto each of the four fundamental subspaces; (*iv*) From the SVD, the (Moore–Penrose) pseudoinverse.

$$a)\quad \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \qquad b)\quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad c)\quad \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \qquad d)\quad \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix}.$$

[8]These matrices have been specially crafted so that their SVD's can be easily computed by hand. For an arbitrary matrix, this will not be the case. You should know how to perform the hand–based calculations using your constructed SVD's as there will likely be a question on the final requiring such a calculation. Note that because three of these matrices have full-rank, you can *check* some of your results by using the full-rank analytic forms of the pseudoinverse and projectors.