ECE 174 Homework # 4 Solutions

1. The two situations correspond to the mutually exclusive cases of a) $b$ is in the range of $A$, and b) nonzero $b$ is not in the range of $A$.

2. Meyer Problem 4.6.7. Let the data pairs be given by $(x_k, y_k), k = 1, \cdots, m = 11$. (E.g., $(x_1, y_1) = (-5, 2)$, etc.). For both the line fit and the quadratic fit, we want to place the problem in the vector–matrix form,

$$y = A\alpha$$

for $A m \times n$, where $n$ is the dimension of the unknown parameter vector $\alpha$, and solve for $\alpha$ in the least-squares sense. For the linear fit we have,

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix} = A\alpha,$$

with $n = 2$. While for the quadratic fit we have,

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix} = A\alpha,$$

with $n = 3$. Note that in both cases we have the data to fill in numerical values of $y$ and $A$. In both cases we have that the matrix $A$ is full rank (you can check in matlab, but it will usually be true for this type of problem). Thus the least squares solution can be determined as,

$$\hat{\alpha} = (A^T A)^{-1} A^T y.$$

The optimal least–squares error has magnitude,

$$\|e\|^2 = \|y - A\hat{\alpha}\|^2 = (y - A\hat{\alpha})^T (y - A\hat{\alpha}),$$

which can be computed for both the quadratic and linear fits. In this case you will find that the quadratic fit provides a much tighter fit to the data. Once you have the parameters at hand, you can perform the prediction by plugging in the new value for $x_k$.

---

$^1$This answer can be easily understood by drawing a picture of the codomain decomposed into the range of $A$ and its orthogonal complement and then drawing various possibilities for the codomain vector $b$. 

1
Questions:

Given measured data \((x_k, y_k), k = 1, \cdots, m\) can you fit the model,

\[ y \approx \alpha_0 + \alpha_3 x^3 + \alpha_9 x^9? \]

Or the model,

\[ y \approx \alpha x + \beta e^x + \gamma \cos(12x^3)? \]

How could you fit the model,

\[ y \approx \alpha e^{\beta x}, \]

for data obeying \(y_k > 0\)?

3. (a) Viewing \(\omega^T\) as a \(1 \times n\) matrix representation of a linear mapping \(\omega^T: \mathbb{R}^n \to \mathbb{R}\) gives an interpretation of \(\omega^T\) as a rank 1 row matrix with an \((n - 1)\)–dimensional nullspace \(N(\omega^T) \subset \mathcal{X} = \mathbb{R}^n\). This is equivalent to saying that the set of all vectors in \(\mathbb{R}^n\) which are orthogonal to the single (i.e. dimension 1) direction lying along \(\omega\) is an \((n - 1)\)–dimensional subspace of \(\mathbb{R}^n\). This set is precisely \(N(\omega^T)\). Let \(x_p\) be a nonzero “particular solution” to the linear inverse problem,

\[ \omega^T x = d. \quad (1) \]

As we know from our studies of linear inverse problems, the set of all solutions to equation (1) is given by the affine subspace

\[ \mathcal{H} = x_p + N(\omega^T). \]

Thus \(\mathcal{H}\) is a translation of the \((n - 1)\)–dimensional subspace \(N(\omega^T)\) and is therefore an \((n - 1)\)–dimensional affine subspace and hence (by definition) is a hyperplane.

(b) The minimum norm solution \(x_0\) to the inverse problem (1) must satisfy the condition that it lies in the range of the adjoint of \(\omega^T\), which is equivalent to the condition

\[ x_0 = \omega \lambda \]

for some scalar parameter \(\lambda\). (I.e., the minimum distance to the hyperplane must be along the direction of \(\omega\).) Combining this condition with (1) yields \(\lambda = \frac{d}{\|\omega\|^2}\), resulting in

\[ x_0 = \frac{d}{\|\omega\|^2} \omega = \frac{d}{\|\omega\|} \frac{\omega}{\|\omega\|} = \frac{d}{\|\omega\|} n \]

---

2 The standard inner product is assumed to hold on all spaces.

3 Namely, vectors in \(N(\omega^T) \subset \mathbb{R}^n\) can have components along all directions but one; they cannot have a nonzero component along the direction \(\omega\). This is because \(x \in N(\omega^T) \iff (\omega, x) = \omega^T x = 0\).

4 Since \(\omega^T\) is onto, a particular solution \(x_p\) is guaranteed to exist.
where \( n = \frac{\omega}{\|\omega\|} \) is the unit vector pointing in the same direction as \( \omega \). The norm of \( x_0 \), \( \|x_0\| = \frac{|d|}{\|\omega\|} \), gives the minimum (unsigned) distance from the origin to the hyperplane \( \mathcal{H} \). The quantity

\[
\Delta \triangleq \frac{d}{\|\omega\|}
\]

gives the minimum signed distance from the origin to the hyperplane, where a positive value for \( d \) indicates that the vector \( x_0 \) points in the same direction as \( \omega \) while a negative value for \( d \) indicates that \( x_0 \) points in the opposite direction to the vector \( \omega \). Note, then, that we have obtained

\[
x_0 = \Delta n,
\]

where \( x_0 \) represents the minimum distance displacement required to move the origin to the hyperplane where this displacement is equal to the (signed) distance \( \Delta \) along the direction given by the unit vector \( n \).

(c) Define

\[
D(x) \triangleq \frac{h(x)}{\|\omega\|} = n^T x - \Delta
\]

and note that \( n \) is a unit vector perpendicular to any displacement vector of the form \( v = x_1 - x_2, x_1, x_2 \in \mathcal{H} = \mathcal{x}_p + \mathcal{N}(\omega^T) \),

\[
n^T v = n^T x_1 - n^T x_2 = d - d = 0.
\]

This shows that \( n \) is perpendicular to the hyperplane \( \mathcal{H} \) (as expected). That \( \Delta \) is a signed distance has been discussed above.

Obviously \( D(x) \) and \( h(x) \) have the same sign for all \( x \in \mathbb{R}^n \) so that the perceptron can be equivalently defined as a hard–thresholding of either function.

(d) Note that any vector \( x \in \mathbb{R}^n \) can be written as

\[
x = x_\parallel + x_\perp
\]

where \( x_\parallel = nn^T x \in \mathcal{R}(\omega) \) is the projection of \( x \) along (parallel to) the direction \( n \) (equivalently, along the direction \( \omega \)) and \( x_\perp = x - x_\parallel = (I - nn^T)x \in \mathcal{R}(\omega)^\perp = \mathcal{N}(\omega^T) \) is the component of \( x \) which is perpendicular to \( n \) (equivalently, perpendicular to \( \omega \)). Note, then, that the matrix \( nn^T \) gives the orthogonal projection onto \( \mathcal{R}(\omega) = \mathcal{N}(\omega^T)^\perp \) while the matrix \( (I - nn^T) \) gives the orthogonal projection onto \( \mathcal{N}(\omega^T) \).

Let the point \( x_{\text{opt}} \in \mathcal{H} \) be the orthogonal projection of an arbitrary point \( x \) onto the hyperplane \( \mathcal{H} \). The length of the vector \( x - x_{\text{opt}} \) gives the minimum distance of the point \( x \) to the hyperplane \( \mathcal{H} \) which corresponds to the requirement that
\(x - x_{\text{opt}} \perp \mathcal{N}(\omega^T)\). This is equivalent to the condition \((I - nn^T)(x - x_{\text{opt}}) = 0\) which can be written as

\[x - x_{\text{opt}} = nn^T(x - x_{\text{opt}}) = n(n^T x - n^T x_{\text{opt}}) = n(n^T x - \Delta) = D(x) n,\]

using the fact that \(x_{\text{opt}} \in \mathcal{H} \Rightarrow n^T x_{\text{opt}} = \Delta\). Thus \(D(x)\) gives the signed distance of the point \(x\) to the hyperplane \(\mathcal{H}\). The minimum (unsigned) distance of \(x\) to \(\mathcal{H}\) is then given by \(\|x - x_0\| = |\Delta|\).

By writing \(D(x)\) as

\[D(x) = n^T x - \Delta = n^T x - n^T x_0 = n^T(x - x_0)\]

we see that \(D(x)\) is the signed length of \(x\) projected along the direction \(n\) and re-referenced to the minimum distance point \(x_0 = \Delta \cdot n\) found above.

Thus \(D(x)\) gives a measure of the degree of penetration of \(x\) into the half-space \(S^+\) (if the signed distance \(D(x)\) is positive) or into the half-space \(S^-\) (if the signed distance \(D(x)\) is negative). Because the hard-limiter function only cares about the sign of \(D(x)\), we see that the perceptron throws away the information about the distance of the point \(x\) to the half-spaces separated by the hyperplane \(\mathcal{H}\).\(^5\)

4. Note that a singular value decomposition (SVD) is not fully unique. However, the singular values are unique. Additional unique quantities are the projection matrices onto the four fundamental subspaces and the pseudoinverse, all of which can be computed from knowledge of the SVD.

(a) We first determine that \(m = 3, n = 1, r = 1, \nu = 0, \) and \(\mu = 2\). Once the dimensions of the various subspaces are known, we can systematically work the matrix \(A\) into the appropriate factorization,\(^6\)

\[
A = \begin{pmatrix}
1 \\
0 \\
1
\end{pmatrix} = \begin{pmatrix}
\frac{1}{\sqrt{2}} \\
0 \\
\frac{1}{\sqrt{2}}
\end{pmatrix} (\sqrt{2})(1) = U_1 S V_1^T
\]

\[
= \begin{pmatrix}
\frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\
0 & 1 & 0 \\
\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}}
\end{pmatrix} (\sqrt{2}) (1) = (U_1 \quad U_2) \begin{pmatrix}
S \\
\emptyset
\end{pmatrix} (V_1^T) = U \Sigma V^T.
\]

All the columns of \(U\) provide an orthonormal basis for \(\mathcal{Y} = \mathbb{R}^3\). The column of \(U_1\) spans \(\mathcal{R}(A)\) and the two columns of \(U_2\) provide an orthonormal basis for \(\mathcal{N}(A^T)\).

\(^5\)I.e., the perceptron cares only about the fact, but not the degree, of penetration into one or another of the half-spaces \(S^-\) and \(S^+\). A probabilistic generalization of the perceptron known as logistic regression exploits all of the information provided by the distance function \(D(x)\).

\(^6\)Note that these are very specially constructed matrices which have been hand crafted to be amenable to this approach. Generally this approach cannot be applied and numerical techniques must be used to obtain the SVD.
The single element of $V^T = V_1^T$ spans $\mathcal{R}(A^T) = \mathbb{R} = \mathcal{X}$. The nullspace of $A$ is trivial. With the SVD in hand, it is readily determined that $\sigma_1 = \sqrt{2} > 0$ and

$$A^+ = \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 \end{pmatrix};$$

$$P_{\mathcal{R}(A)} = \frac{1}{2} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix};$$

$$P_{\mathcal{N}(A^T)} = \frac{1}{2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 1 \end{pmatrix};$$

$$P_{\mathcal{R}(A^T)} = 1;$$

$$P_{\mathcal{N}(A)} = 0.$$

Note that because $A$ has full column rank we can also compute $A^+$ directly as $A^+ = (A^T A)^{-1} A^T$.

(b) We first determine that $m = 3$, $n = 2$, $r = 2$, $\nu = 0$, and $\mu = 1$. Again, once the dimensions of the various subspaces are known, we can systematically work the matrix $A$ into the appropriate factorization,

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \\ 0 & \sqrt{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 \\ 0 & 1 \\ \frac{1}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \\ 0 & \sqrt{2} \end{pmatrix} = U_1 S V_1^T$$

$$= \begin{pmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} = (U_1 \ U_2) \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix} = U \Sigma V^T.$$

All the columns of $U$ provide an orthonormal basis for $\mathcal{Y} = \mathbb{R}^3$. The two columns of $U_1$ give an orthonormal basis for $\mathcal{R}(A)$ and the column of $U_2$ spans $\mathcal{N}(A^T)$. Both rows of $V^T = V_1^T$ provide an orthonormal basis for $\mathcal{R}(A^T) = \mathbb{R}^2 = \mathcal{X}$. The nullspace of $A$ is trivial. It now can be readily determined that $\sigma_1 = \frac{\sqrt{2}}{2} > \sigma_2 = 1 > 0$ and

$$A^+ = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix};$$

$$P_{\mathcal{R}(A)} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix};$$

$$P_{\mathcal{N}(A^T)} = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix};$$

$$P_{\mathcal{R}(A)} = I, \quad P_{\mathcal{N}(A)} = 0.$$
Note that because \( A \) has full column rank we can also compute \( A^+ \) directly as \( A^+ = (A^T A)^{-1} A^T \).

(c) Here \( m = n = 3 \) while \( r = 2 \). For this case, the matrix \( A \) is rank-deficient (i.e., is neither one-to-one nor onto) and hence the pseudoinverse, \( A^+ \), cannot be constructed using an analytic formula as can be done for the full-rank cases. We also have \( \nu = \mu = 1 \). The matrix \( A \) factors as

\[
A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} = \left( \begin{array}{c} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{array} \right) \left( \begin{array}{ccc} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 & \sqrt{2} \\ \frac{1}{\sqrt{2}} & \sqrt{2} & 0 \end{array} \right) \left( \begin{array}{ccc} 1 & 0 & 1 \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{array} \right) = U_1 S V_1^T
\]

All the columns of \( U \) provide an orthonormal basis for \( \mathcal{Y} = \mathbb{R}^3 \). The two columns of \( U_1 \) gives an orthonormal basis for \( \mathcal{R}(A) \) and the column of \( U_2 \) spans \( \mathcal{N}(A^T) \). All the rows of \( V^T \) provide an orthonormal basis for \( \mathcal{X} = \mathbb{R}^3 \). The two rows of \( V_1^T \) provide an orthonormal basis for \( \mathcal{R}(A^T) \). The nullspace of \( A \) is spanned by the row of \( V_2^T \). It can now be readily determined that \( \sigma_1 = 2 > \sigma_2 = 1 > 0 \) and

\[
P_{\mathcal{R}(A)} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix};
\]

\[
P_{\mathcal{N}(A^T)} = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix};
\]

\[
P_{\mathcal{R}(A^T)} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix};
\]

\[
P_{\mathcal{N}(A)} = \begin{pmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.
\]
Note that because $A$ is symmetric it just happens to be the case that $P_{\mathcal{R}(A)} = P_{\mathcal{R}(A^T)}$ and $P_{\mathcal{N}(A)} = P_{\mathcal{N}(A^T)}$, however this is not true for a general nonsymmetric matrix $A$.

(d) Here $m = 2$, $n = 3$, and $r = 1$. Thus $A$ is rank-deficient and non-square (and definitely nonsymmetric).

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{pmatrix}$$
$$= \begin{pmatrix} 1 \\ 2 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} (\sqrt{5}) \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{pmatrix} (\sqrt{15}) \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix} = U_1 S V_1^T$$
$$= \begin{pmatrix} \frac{1}{\sqrt{5}} \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{15} \\ 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{\sqrt{5}} \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{15} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{3}} \frac{1}{\sqrt{3}} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \frac{1}{\sqrt{3}} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} \frac{1}{\sqrt{6}} \frac{1}{\sqrt{6}} \end{pmatrix}$$
$$= U \Sigma V^T.$$

Both columns of $U$ provide an orthonormal basis for $\mathcal{Y} = \mathbb{R}^2$. The column of $U_1$ spans $\mathcal{R}(A)$ and the column of $U_2$ spans $\mathcal{N}(A^T)$. All the rows of $V^T$ provide an orthonormal basis for $\mathcal{X} = \mathbb{R}^3$. The row of $V_1^T$ spans $\mathcal{R}(A^T)$. The nullspace of $A$ is spanned by the two rows of $V_2^T$. It can now be readily determined that

---

7The special properties $P_{\mathcal{R}(A)} = P_{\mathcal{R}(A^*)}$ and $P_{\mathcal{N}(A)} = P_{\mathcal{N}(A^*)}$ hold for any self-adjoint matrix $A$, $A = A^*$. 

---
\( \sigma_1 = \sqrt{15} > 0 \) and

\[
A^+ = \frac{1}{15} \begin{pmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{pmatrix};
\]

\[
P_{R(A)} = \frac{1}{5} \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix};
\]

\[
P_{N(A^T)} = \frac{1}{5} \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix};
\]

\[
P_{R(A^T)} = \frac{1}{3} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix};
\]

\[
P_{N(A)} = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.
\]