Problem 2.4 *worth 4 points*
Linear function? The function $\phi : \mathbf{R}^3 \to \mathbf{R}$ satisfies

$$\phi(1,1,0) = -1 \quad \phi(-1,1,1) = 1 \quad \phi(1,-1,-1) = 1$$

Choose one of the following, and justify your choice: ϕ must be linear; ϕ could be linear; ϕ cannot be linear.

Solution:

$$\alpha \cdot \phi(\vec{x}) = \phi(\alpha \cdot \vec{x}) \quad (\textit{principle of superposition})$$

Let $\alpha = -1$, $\vec{x}_b = (-1,1,1)$, and $\vec{x}_c = (1,-1,-1)$

$$\begin{aligned} \alpha \cdot \phi(\vec{x}_b) &= \phi(\alpha \cdot \vec{x}_b) \\ -1 \cdot \phi(-1,1,1) &= \phi(-1 \cdot \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}) \\ -1 \cdot (1) &= \phi(\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}) \\ -1 &= \phi(1,-1,-1) \\ -1 &\neq 1 \end{aligned}$$

Because we are told the function $\phi(\vec{x}_c) = 1$ when $\vec{x}_c = (1,-1,-1)$, and since $\alpha \cdot \vec{x}_b = (1,-1,-1) = \vec{x}_c$, we see that superposition does not hold, therefore ϕ cannot be linear.

Problem 2.4 *worth 4 points*
Affine function. Suppose $\psi : \mathbf{R}^2 \to \mathbf{R}$ is an affine function, with $\psi(1,0) = 1$, $\psi(1,-2) = 2$.

(a) What can you say about $\psi(1,-1)$? Either give the value of $\psi(1,-1)$ or state that it cannot be determined.

(b) What can you say about $\psi(2,-2)$? Either give the value of $\psi(2,-2)$ or state that it cannot be determined.

Solution:

(a) Since we're given $\psi(1,0) = 1$ and $\psi(1,-2) = 2$ we can conclude that the value $1 < \psi(1,-1) < 2$ because we see $\vec{x}_1$ remains constant as $\vec{x}_2$ is changing which, in turn, causes $\psi(\vec{x})$ to change. Since an affine function must be linear, if we keep a value constant and change another the result will change linearly, and since $|\Delta\vec{x}_2| = 2$ and $|\Delta\psi(\vec{x})| = 1$, if we look at the change in result with respect to $\vec{x}_2$ we get $\frac{1}{2}$. Add this on to our given initial value and we get $\psi(1,-1) = \frac{3}{2}$.

(b) We are not told how the function behaves as $\vec{x}_1$ changes, so the value of $\psi(2,-2)$ cannot be determined.

Problem 2.9 *worth 6 points*
Taylor approximation. Consider the function $f : \mathbf{R}^2 \to \mathbf{R}$ given by $f(x_1, x_2) = x_1 x_2$. Find the Taylor approximation $\hat{f}$ at the point $z = (1, 1)$. Compare $f(x)$ and $\hat{f}(x)$ for the following values of x.

$$x = (1, 1), \quad x = (1.05, 0.95), \quad x = (0.85, 1.25), \quad x = (-1, 2).$$

Make a brief comment about the accuracy of the Taylor approximation in each case.

Solution: If $z = (1, 1)$, then we see $f(z) = (1) \cdot (1) = 1$.

$$\hat{f}(x) = f(z) + \frac{\delta f(z)}{\delta x_1}(x_1 - z_1) + \frac{\delta f(z)}{\delta x_2}(x_2 - z_2)$$

Now we take a look at our partial derivatives for x_1 and x_2 when $f(z) = 1$

$$\begin{aligned} \frac{\delta f(z)}{\delta x_1} &= x_2(1) \\ &= (1)(1) \\ &= 1 \end{aligned} \qquad\qquad \begin{aligned} \frac{\delta f(z)}{\delta x_2} &= x_1(1) \\ &= (1)(1) \\ &= 1 \end{aligned}$$

Going back to our function $\hat{f}$

$$\begin{aligned} \hat{f}(x) &= f(z) + \frac{\delta f(z)}{\delta x_1}(x_1 - z_1) + \frac{\delta f(z)}{\delta x_2}(x_2 - z_2) \\ \hat{f}(x) &= (1) + (1)(x_1 - 1) + (1)(x_2 - 1) \\ \hat{f}(x) &= (1) + (x_1 - 1) + (x_2 - 1) \end{aligned}$$

Now let's evaluate for each of our given vectors:

$$\begin{aligned} \hat{f}(1, 1) &= (1) + (1 - 1) + (1 - 1) \\ \hat{f}(1, 1) &= 1 + 0 + 0 \\ \hat{f}(1, 1) &= 1 \end{aligned}$$

$$\begin{aligned} \hat{f}(1.05, 0.95) &= (1) + (1.05 - 1) + (.95 - 1) \\ \hat{f}(1.05, 0.95) &= (1) + (.05) - (.05) \\ \hat{f}(1.05, 0.95) &= 1 \end{aligned}$$

$$\begin{aligned} \hat{f}(0.85, 1.25) &= (1) + (.85 - 1) + (1.25 - 1) \\ \hat{f}(0.85, 1.25) &= (1) - (.15) + (.25) \\ \hat{f}(0.85, 1.25) &= 1.1 \end{aligned}$$

$$\begin{aligned} \hat{f}(-1, 2) &= (1) + (-1 - 1) + (2 - 1) \\ \hat{f}(-1, 2) &= (1) - 2 + 1 \\ \hat{f}(-1, 2) &= 0 \end{aligned}$$

What each of these numbers says about the point is an estimation for the value of $f(x)$, based on the value of $f(z)$ and point x's proximity to z. For vectors $x_1 = (1, 1)$ and $x_2 = (1.05, 0.95)$ we can see that our values should be close to one ($f(x_1) = 1$ and $f(x_2) = .9975$ in these cases) so we get a value of 1. These are very accurate because if we look at our points as edges of a triangle then using Pythagorean formula ($a^2 + b^2 = c^2$ where a is given by Δx_1 and b is given by Δx_2) we see the following values for each of our points:

$$\begin{aligned}
(1,1) &= \sqrt{(1-1)^2 + (1-1)^2} \\
(1,1) &= \sqrt{0} \\
(1,1) &= 0 && \textit{(0 because we're at the same point)}
\end{aligned}$$

$$\begin{aligned}
(1.05, 0.95) &= \sqrt{(1-1.05)^2 + (1-0.95)^2} \\
(1.05, 0.95) &= \sqrt{.0025 + .0025} \\
(1.05, 0.95) &= .0707 && \textit{(.07 says we're very close)}
\end{aligned}$$

$$\begin{aligned}
(0.85, 1.25) &= \sqrt{(1-0.85)^2 + (1-1.25)^2} \\
(0.85, 1.25) &= \sqrt{.0225 + .0625} \\
(0.85, 1.25) &= .29 && \textit{(.29 says we're further)}
\end{aligned}$$

$$\begin{aligned}
(-1, 2) &= \sqrt{(1-(-1))^2 + (1-2)^2} \\
(-1, 2) &= \sqrt{4+1} \\
(-1, 2) &= 2.236 && \textit{(2.236 says we're very far from our point)}
\end{aligned}$$

For vector $x_3 = (0.85, 1.25)$, our function should return the value 1.0625 for x_3, and we see a rounded version of this value as our approximation. For $x_4 = (-1, 2)$, we see our point is much further away than our other 3 as they compare to z. Our function $f(x_4)$ would return the value -2, but since we are approximating a value based on our current known value for z, we see the accuracy decreases the further away we are, and we get 0 as our answer.

Problem 2.10 *worth 6 points*
Regression model. Consider the regression model $\hat{y} = x^T\beta + v$, where $\hat{y}$ is the predicted response, x is an 8-vector of features, β is an 8-vector of coefficients, and v is the offset term. Determine whether each of the following statements is true or false.

(a) If $\beta_3 > 0$, and $x_3 > 0$, then $\hat{y} \geq 0$.

(b) If $\beta_2 = 0$ then the prediction $\hat{y}$ does not depend on the second feature x_2.

(c) If $\beta_6 = -0.8$, then increasing x_6 (keeping all other x_is the same) will decrease $\hat{y}$.

Solution:

(a) False. We can't conclude the value of $\hat{y}$ from just a singular x_i when we have a vector of 8 features.

(b) True. If $\beta_2 = 0$ then it "eliminates" the value at x_2 meaning the final product does not depend on that value.

(c) True. Increasing the value of x_6 will create a larger negative value. Since $x^T\beta$ involves summing all our elements after multiplying them with each corresponding coefficient, and since we are leaving all other x_is the same, then increasing our value x_6 will create a larger negative value in the summation, creating a smaller scalar of $\hat{y}$.