DSC 140A - Homework 02

Due: Wednesday, January 24

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope at 11:59 PM.

Problem 1.

Let A be a symmetric $n \times n$ matrix, and let $\vec{x} \in \mathbb{R}^n$. Let $f(\vec{x}) = \vec{x}^T A \vec{x}$. Show that

$$\frac{df}{d\vec{x}} = 2A\vec{x}.$$

Hint: this problem was started in discussion.

Solution: First we rewrite $A\vec{x}$ as it's sum of components:

$$A\vec{x} = \sum_{j=1}^{n} x_j \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}$$

Then we can rewrite $f(\vec{x})$ as:

$$f(\vec{x}) = \vec{x}^T A \vec{x} = \sum_{i=1}^{n} x_i \sum_{j=1}^{n} x_j a_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij}$$

Now we can take the partial derivative of $f(\vec{x})$ with respect to x_k:

$$\frac{\partial f}{\partial x_k} = \frac{\partial}{\partial x_k} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij}$$

Looking at when $k = 1$:

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j a_{ij}$$

There are 2 cases to consider, when $i = 1$, when $j = 1$, looking at the first case:

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1} \sum_{j=1}^{n} x_1 x_j a_{1j} = \sum_{j=1}^{n} x_j a_{1j}$$

Looking at the second case:

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1} \sum_{i=1}^{n} x_i x_1 a_{i1} = \sum_{i=1}^{n} x_i a_{i1}$$

Combining the two cases:

$$\frac{\partial f}{\partial x_1} = \sum_{j=1}^{n} x_j a_{1j} + \sum_{i=1}^{n} x_i a_{i1}$$

And since A is symmetric, $a_{ij} = a_{ji}$, so we can rewrite the above as:

$$\frac{\partial f}{\partial x_1} = 2\sum_{j=1}^{n} x_j a_{1j}$$

So we can generalize this to all k:

$$\frac{\partial f}{\partial x_k} = 2\sum_{j=1}^{n} x_j a_{kj}$$

And since this is true for all k, we can rewrite this as:

$$\frac{\partial f}{\partial \vec{x}} = \begin{pmatrix} 2\sum_{j=1}^{n} x_j a_{1j} \\ 2\sum_{j=1}^{n} x_j a_{2j} \\ \vdots \\ 2\sum_{j=1}^{n} x_j a_{nj} \end{pmatrix} = \begin{pmatrix} 2a_{11} & 2a_{12} & \cdots & 2a_{1n} \\ 2a_{21} & 2a_{22} & \cdots & 2a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 2a_{n1} & 2a_{n2} & \cdots & 2a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = 2A\vec{x}$$

Problem 2.

Consider a linear prediction function H used for binary classification, and assume that when the output of H is positive we predict for class +1, and when it's negative we predict for class -1. This means that the **decision boundary** is where $H(\vec{x}) = 0$.

In lecture, we saw that a linear prediction function has the form:

$$\begin{aligned} H(\vec{x}) &= w_0 + w_1 x_1 + \ldots + w_d x_d \\ &= \vec{w} \cdot \text{Aug}(\vec{x}) \end{aligned}$$

where $\vec{w} = (w_0, w_1, \ldots, w_d)^T$. In this problem, it will also be useful to define the vector $\vec{w}' = (w_1, \ldots, w_d)^T$, which is the same as $\vec{w}$ except that it does not include w_0. Note that $\vec{x} \cdot \vec{w}' = w_1 x_1 + \ldots w_d x_d$. With this definition, we can write $H(\vec{x})$ in a slightly different way:

$$H(\vec{x}) = w_0 + \vec{w}' \cdot \vec{x}$$

Remember this formula, as it will be useful several times below!

Over the course of this problem, we'll answer the question: how is the magnitude of $H(\vec{x})$ related to the distance between $\vec{x}$ and the decision boundary?

Note: for this problem it may be useful to review the properties of the dot product and vector algebra. In particular, remember that $\|u\|$ denotes the norm (length) of a vector, and that $\vec{u} \cdot \vec{u} = \|u\|^2$. By dividing a vector by its norm, as in $\vec{u}/\|\vec{u}\|$, we obtain a *unit vector* with unit length – a unit vector is useful for specifying a direction. To find the component of a vector $\vec{u}$ that points in the same direction as another vector $\vec{v}$, we write $\vec{u} \cdot \vec{v}/\|\vec{v}\|$. Also remember that $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$, and that $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$.

a) Show that for any point $\vec{z}$ on the decision boundary, $\vec{w}' \cdot \vec{z} = -w_0$.

Hint: use that fact that $H(\vec{z}) = 0$.

Solution: We can use that fact that $H(\vec{z}) = 0$, then we can rewrite this as:

$$H(\vec{z}) = w_0 + \vec{w}' \cdot \vec{z} = 0$$

Then rearranging this equation we get:

$$\vec{w}' \cdot \vec{z} = -w_0$$

As desired.

b) Argue that $\vec{w}'$ is orthogonal to the decision boundary.

Hint: take two arbitrary points $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ that are assumed to be on the decision boundary. Then, since we know that the boundary is linear (it is a line, plane, etc.), the difference of these vectors, $\vec{\delta} = \vec{x}^{(1)} - \vec{x}^{(2)}$ is parallel to the boundary. To show that $\vec{w}'$ is orthogonal to the boundary, it suffices to show that $\vec{w}' \cdot \vec{\delta} = 0$. Make sure you somehow use the fact that $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ are on the decision boundary.

Solution: As said in the question, let $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ be two arbitrary points on the decision boundary. Then we can write:

$$\vec{w}' \cdot \vec{x}^{(1)} + \vec{w}_0 = 0$$
$$\vec{w}' \cdot \vec{x}^{(2)} + \vec{w}_0 = 0$$

And we can also definre $\vec{\delta} = \vec{x}^{(1)} - \vec{x}^{(2)}$. Then subtracting the two equations above we get:

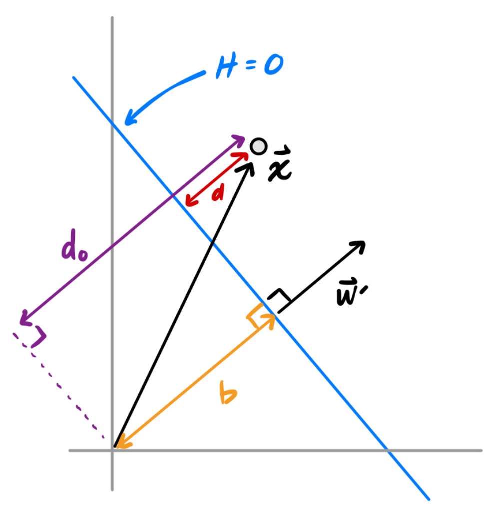
$$\vec{w}'(\vec{x}^{(1)} - \vec{x}^{(2)}) = 0$$

And $\delta = \vec{x}^{(1)} - \vec{x}^{(2)}$, thus,

$$\vec{w}' \cdot \vec{\delta} = 0$$

And since $\vec{x}^{(1)}$ and $\vec{x}^{(2)}$ are on the decision boundary, it means that they are parallel to the boundary, and since $\vec{\delta}$ is the difference of the two points, it is also parallel to the boundary. Thus, $\vec{w}'$ is orthogonal to the decision boundary.

c) Now that we know that $\vec{w}'$ is orthogonal to the decision boundary, we can draw a better picture of the situation:



The blue line is the decision boundary – it is where $H = 0$. We have drawn an arbitary point $\vec{x}$, along with several distances:

- d: the (signed) distance from the decision boundary to $\vec{x}$

- b: the distance from the the origin to the decision boundary
- d_0: the length of the component of $\vec{x}$ that is orthogonal to the decision boundary.

We're most interested in knowing d. First, though, we need to find b.

Consider the vector $b\vec{w}'/\|\vec{w}'\|$; this is a vector from the origin to the decision boundary that is orthogonal to the boundary and with length b.

Since this vector is on the decision boundary, $H(b\vec{w}'/\|\vec{w}'\|) = 0$. Using this fact, show that $b = -\frac{w_0}{\|\vec{w}'\|}$.

Hint: first show that $H(b\vec{w}'/\|\vec{w}'\|) = b\|\vec{w}'\| + w_0$, then set this to zero and solve for b.

Solution: Using $H(b\vec{w}'/\|\vec{w}'\|) = 0$, we can rewrite this as:

$$\begin{aligned} H(b\vec{w}'/\|\vec{w}'\|) &= b\frac{\vec{w}' \cdot \vec{w}'}{\|\vec{w}'\|} + w_0 \\ b\|\vec{w}'\| + w_0 &= 0 \\ b\|\vec{w}'\| &= -w_0 \\ b &= -\frac{w_0}{\|\vec{w}'\|} \end{aligned}$$

d) Recall that d_0 is the component of $\vec{x}$ that is orthogonal to the decision boundary; this is simply $\vec{x} \cdot \vec{w}'/\|\vec{w}'\|$. From the picture, $d_0 = d + b$. We know d_0 and b, and can therefore solve for d.

Use this to show that $|d| = |H(\vec{x})|/\|\vec{w}'\|$.

Solution: Using all the information above, we can write:

$$\begin{aligned} d &= d_0 - b \\ &= \frac{\vec{x} \cdot \vec{w}'}{\|\vec{w}'\|} + \frac{\vec{w}_0}{\|\vec{w}'\|} \\ &= \frac{\vec{x} \cdot \vec{w}' + \vec{w}_0}{\|\vec{w}'\|} \\ &= \frac{H(\vec{x})}{\|\vec{w}'\|}, \quad \text{since } H(\vec{x}) = \vec{x} \cdot \vec{w}' + \vec{w}_0 \end{aligned}$$

And since d is the distance from the decision boundary to $\vec{x}$, we can write $|d| = |H(\vec{x})|/\|\vec{w}'\|$. Getting the result we wanted.

We have shown that the distance between $\vec{x}$ and the decision boundary is proportional to the output of the prediction function, $H(\vec{x})$. This gives us a very useful interpretation of $|H(\vec{x})|$! For example, this means if $H(\vec{x}^{(1)}) > H(\vec{x}^{(2)}) > 0$, then $\vec{x}^{(1)}$ is further from the decision boundary than $\vec{x}^{(2)}$.

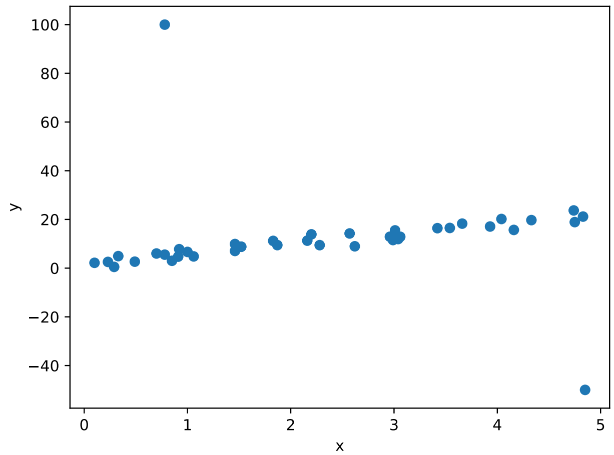
Problem 3.

The file at the link below contains a simple data set suitable for regression.

`https://f000.backblazeb2.com/file/jeldridge-data/002-regression_outlier/data.csv`

The first column contains the x values (the predictor variable) and the second column contais the y values (the target).

The plot below shows the data:



Notably, the data contains outliers which may affect our regression.

Fit a linear function of the form $w_0 + w_1 x$ by minimizing the mean squared error. Report w_0 and w_1, and include your code.

Hint: You may use whatever language or tool you'd like, but if you're using Python, take a look at `np.linalg.lstsq`.

Code

```python
import numpy as np
import matplotlib.pyplot as plt

data = np.loadtxt('data.csv', delimiter=',')

data = np.c_[np.ones(data.shape[0]), data]

w0, w1 =  np.linalg.lstsq(data[:,[0,1]], data[:,2])[0]

print('w0 = {}, w1 = {}'.format(w0, w1))
```

Solution: The values are $w_0 = 11.259$ and $w_1 = 0.180$ rounded to three decimal places.