Sheet 3

Thomas Lawson

November 3, 2025

Questions for Andrew

- Explain to solve Question 3 Part IV. $(\omega_1, \omega_2)^T$ is a point, not a linear equation so unsure how to verify that it is orthogonal to the decision boundary
-

Question 1

Consider a two class problem where a single feature is being observed in order to perform classification. Some training data is sampled and hand-labeled. The data for the two classes are as follows

- ω_1 : 108 88 93 112 99
- ω_2 : 148 152 128 133 139

I. **Training**: Assuming the observations for each class to be normally distributed, estimate the parameters of a Bayesian classifier

Here is the equation for a normal distribution:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ represents the mid point of the distribution i.e. the mean. We can easily calculate this by finding the mean of each of the two datasets.

$$\mu_1 = \frac{1}{5}\sum_{5}^{i=1} x_{1i} = \frac{108+88+93+112+99}{5} = \frac{108+200+192}{5} = \frac{500}{5} = 100$$

$$\mu_2 = \frac{1}{5}\sum_{5}^{i=1} x_{2i} = \frac{148+152+128+133+139}{5} = \frac{300+300+100}{5} = \frac{700}{5} = 20 \times 7 = 140$$

σ^2 is the variance but σ is the standard deviation. I will calculate the variance for each (as I already know the equation for that) and then caclulate the square root.

$$\sigma_1^2 = \frac{1}{5-1}\sum_{5}^{i=1}(x_{1i} - 100)^2 = \frac{64+144+49+144+1}{5} = \frac{288+50+64}{5} = \frac{402}{5} = 80.4$$

$$\sigma_2^2 = \frac{1}{5-1}\sum_{5}^{i=1}(x_{2i} - 100)^2 = \frac{64+144+144+49+1}{5} = \frac{402}{5} = 80.4$$

Now we know the values for standard deviation are $\sqrt{80.4}$ for both σ_1 and σ_2

II. **Testing**: A set of 3 unknown objects is observed. The observations are 115, 118 and 121. Classify each object.

If we were to plot these points on a 1-D graph, we would get no overlap between ω_1 and ω_2. I will set the decision boundry as halfway between the highest ω_1 value and the lowest ω_2 value.
$\frac{112+128}{2} = \frac{240}{2} = 120$

115 is smaller than 120 $\therefore$ we'd label it as ω_1

118 is smaller than 120 $\therefore$ we'd label it as ω_1

121 is greater than 120 $\therefore$ we'd label it as ω_2

III. **Decision boundary**: What is the effective threshold x_0 that distinguishes between the two classes

See part ii

IV. **Classifier evaluation**: If 1,000 randomly-sampled objects are classified how many errors would you expect the classifier to make?

We calculate the z-score for both distributions

$$z_1 = \frac{x_0 - \mu_1}{\sigma_1} = \frac{120 - 100}{\sqrt{80.4}} \approx 2.22$$
$$z_2 = \frac{\mu_2 - x_0}{\sigma_2} = \frac{140 - 120}{\sqrt{80.4}} \approx 2.22$$

From a normal distrubution table we find that a z score of 2.2 gives a 1.3% chance of an error.

Assuming both are classes are equally as likely, here is the probability of an error

$$P(error) = 500 \times 0.013 + 500 \times 0.013 = 6.5 + 6.5 = 13$$

V. **Risk**: If 1 represents healthy cells and 2 represents cancerous cells and the classifier is being used as the 1st stage of a medical screening program. Would you expect the threshold x0 to be used? If not, why not and would the appropriate value be higher or lower?

Always lean towards cancerous cells as the risk associated with labelling an unhealthy cell and healthy has significantly worse consequences for the patient than if it were the other way around.

Question 2 *Linear Classifier - Testing*

I. Use the classifier to group the grid of 9 points, (1,1),(1,2),(1,3), (2,1), (2,2), (2,3), (3,1), (3,2) and (3,3) into two classes.

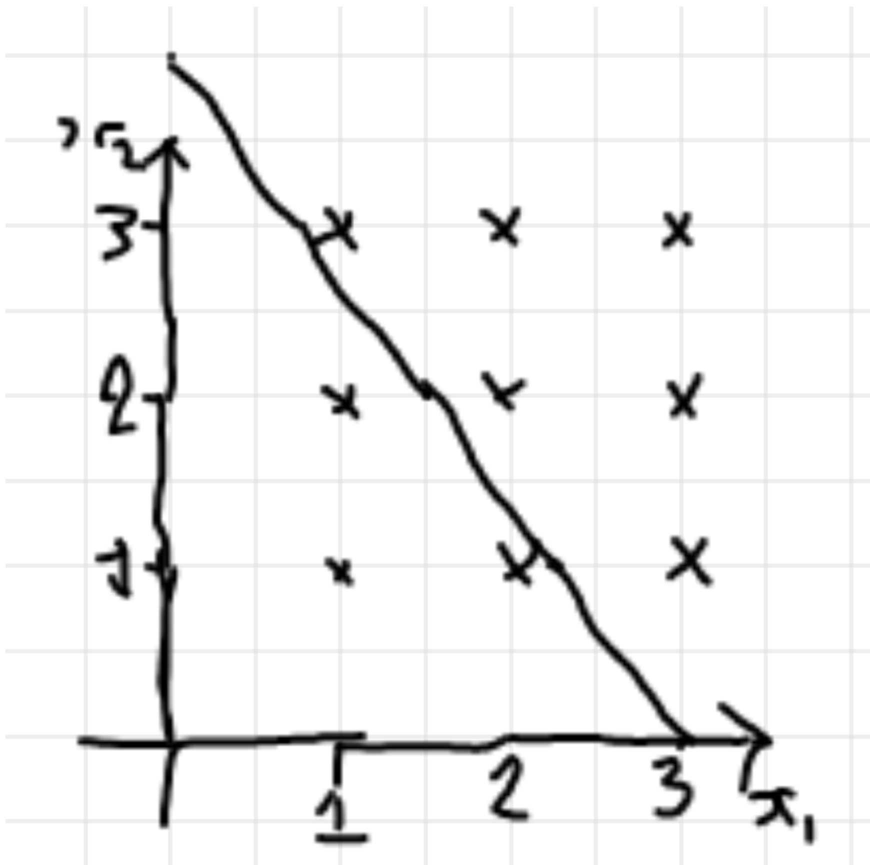
We need to find values for x where $(x_1 \times 4) + (x_2 \times 3) = 12$

$(x_1 \times \omega 1) + (x_2 \times \omega 2) = 12$ when $x = (3, 0)^T$

$(x_1 \times \omega 1) + (x_2 \times \omega 2) = 12$ when $x = (0, 4)^T$

I'll find a third point midway to make it easier to draw on the sketch.

$(x_1 \times \omega 1) + (x_2 \times \omega 2) = 12$ when $x = (1.5, 2)^T$



$(2, 1)\,(1, 2),\text{and}(1, 1)$ would be labelled as ω_1 according to the decision boundary and the rest would be classified as ω_2.

Notes: Didn't realise $\omega^T x$ would be the dot product and just multiplied the two instead originally

II. In the form $x_2 = mx_1 + c$ what is the equation of the decision boundary?

$m = \frac{\delta y}{\delta x} = \frac{4-0}{0-3} = -\frac{4}{3}$

This gives us the equation $x_2 = -\frac{4}{3} + 4$

III. Sketch the decision boundary.

[See part I]

IV. Verify that the decision boundary is orthogonal to $(\omega_1, \omega_2)^T$

Unsure how to answer

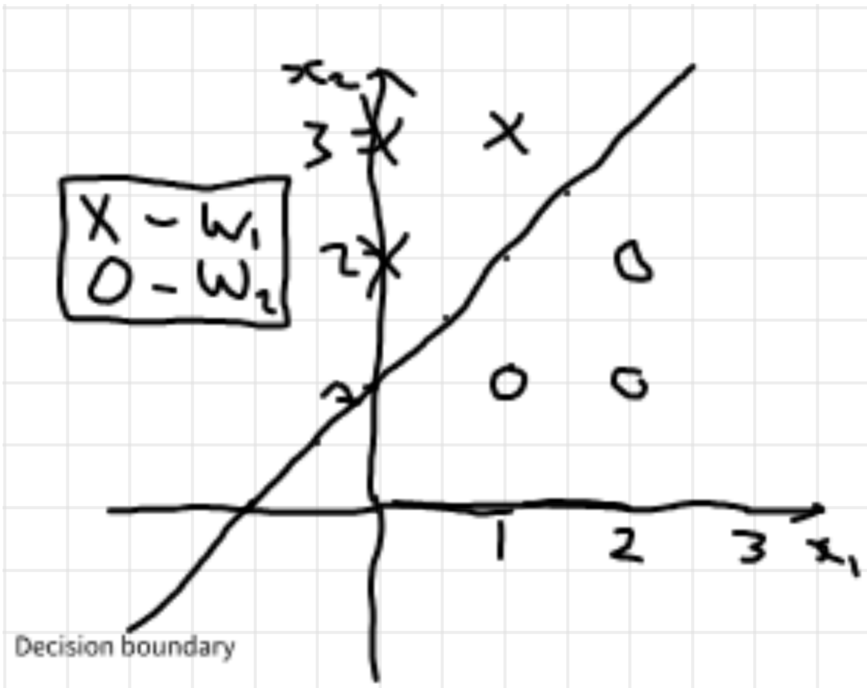
V. If $x_1 = 1.5$ what value of x_2 would lie on the threshold between the two classes?

x_2 would hold the value 2 as calculated in part I.

Question 3 *Linear Classifier - Testing*

Some labeled training has been collected. Observations for class ω_1 are (0,2), (0, 3) and (1,3) and for class ω_2 are (1,1) , (2,1) and (2,2).

I. Plot the data in feature space.



II. Using your sketch find a suitable linear decision boundary.

[See part I]

III. What is the equation of the decision boundary.

The decision boundary is linear $\therefore$ it will take the form $mx + c$.

When $x_1 = 0$, $x_2 = 1$ $\therefore$ $c = 1$

When x_1 increases by 1, so does x_2 $\therefore$ $m = 1$

This gives us the equation for the decision boundary $x_2 = x_1 + 1$.

IV. Find parameters ω_0, ω_1 and ω_2 for the linear classifier that produces this boundary.

A linear classifier has the form:

$$\omega_1 x_1 + \omega_2 x_2 + \omega_0 = 0$$

Compare with the line equation in slope-intercept form:

$$x_2 = -\frac{\omega_1}{\omega_2} x_1 - \frac{\omega_0}{\omega_2}$$

From $x_2 = x_1 + 1$, we can identify:

$$-\frac{\omega_1}{\omega_2} = 1 \quad \Rightarrow \quad \omega_1 = -\omega_2$$

$$-\frac{\omega_0}{\omega_2} = 1 \quad \Rightarrow \quad \omega_0 = -\omega_2$$

Choose a convenient scale for $\omega_2 = 1$, then:

$$\omega_1 = -1, \quad \omega_2 = 1, \quad \omega_0 = -1$$

Therefore, the linear classifier parameters are:

$$\boxed{\omega_0 = -1,\ \ \omega_1 = -1,\ \ \omega_2 = 1}$$

Question 4

For each of the following cases run the perceptron learning algorithm by hand to compute the parameters of the decision boundary. Show your workings. Plot the points and the initial and final decision boundaries.

i)
- data, $\omega_1 : \{(0,1)^T, (1,0)^T\}$, $\omega_2 : \{(1,3)^T, (3,0)^T\}$,
- initial parameters, $w_1 = 1$, $w_2 = -1$, $w_0 = 0$,
- learning rate, 0.5.

First calculate the weighted sums

$$z_1 = 1\times$$

ii)
- data, $\omega_1 : \{(0,0)^T, (1,0)^T, (0,1)^T\}$, $\omega_2 : \{(1,2)^T, (2,1)^T, (2,2)^T\}$,
- initial parameters, $w_1 = 1$, $w_2 = -1$, $w_0 = 0$,
- learning rate, 0.5.

iii)
- data, $\omega_1 : \{(0,0)^T, (1,0)^T, (0,1)^T\}$, $\omega_2 : \{(1,2)^T, (2,1)^T, (2,2)^T\}$,
- initial parameters, $w_1 = 1$, $w_2 = -1$, $w_0 = 0$,
- learning rate, 1.0.

iv)
- data, $\omega_1 : \{(0,0)^T, (1,0)^T, (0,1)^T\}$, $\omega_2 : \{(1,2)^T, (2,1)^T, (2,2)^T\}$,
- initial parameters, $w_1 = 1$, $w_2 = -1$, $w_0 = 0$,
- learning rate, 0.1.