1 (a)

D(a, b) ≥ D(a, c) + D(c, b) 2

(b)
D1 = 1.9 1
D2 = 2.0 1
D3 = 2.1 1
Classify as 1, since D1 is smallest distance hence nearest neighbour 1

(c)
D1 = (2.1 - 2) + (5 - 2.7) = 2.4 1
D2 = (5 - 2.1) + (3 - 2.7) = 3.2 1
D3 = (3 - 2.1) + (2.7 -1) = 2.6 1
Hence classify as class 1 since D1 is the smallest distance. 1

10

also accept x $\frac{1}{2}$

2(a)

$$\lambda(\alpha_1|\omega_1) = 2$$

$$\lambda(\alpha_2|\omega_1) = 3$$

$$\lambda(\alpha_1|\omega_2) = 32$$

$$\lambda(\alpha_2|\omega_2) = 18$$

(b)

$$R(\alpha_1) = \lambda(\alpha_1|\omega_1)P(\omega_1|\alpha_1) + \lambda(\alpha_1|\omega_2)P(\omega_2|\alpha_1) = 2 \times 0.93 + 32 \times 0.07 = 4.1$$

$$R(\alpha_2) = \lambda(\alpha_2|\omega_1)P(\omega_1|\alpha_2) + \lambda(\alpha_2|\omega_2)P(\omega_2|\alpha_2) = 3 \times 0.95 + 18 \times 0.05 = 3.75$$

Hence select from company B. 2

12

3 $P(\omega_n|x) = P(x|\omega_n)P(\omega_n)/P(x)$

$P(x|\omega_n)P(\omega_n) = 0.06, 0.08, 0.12$ for n = 1, 2, 3 respectively

Sum these to get $P(x) = 0.26$

$$P(\omega_1|x) = 0.231$$

$$P(\omega_2|x) = 0.308$$

$$P(\omega_3|x) = 0.462$$

Hence pick ω_3 since maximum posterior probability.

6

4 (a) Take two points a and b on the surface:

$$\boldsymbol{w}^t\boldsymbol{a} + w_0 = \boldsymbol{w}^t\boldsymbol{b} + w_0$$

$$\boldsymbol{w}^t\boldsymbol{a} = \boldsymbol{w}^t\boldsymbol{b}$$ 2

$$\boldsymbol{w}^t(\boldsymbol{a} - \boldsymbol{b}) = 0$$

Vector $\boldsymbol{a} - \boldsymbol{b}$ must lie parallel to the surface. $\boldsymbol{w}^t$ must be normal to $\boldsymbol{a} - \boldsymbol{b}$ so that the dot product = 0 as above. Therefore $\boldsymbol{w}^t$ must be normal to the surface. 2

(b) $10 + 2 \times -3 + \ -7 \times 2 + 1 \times 9 = -1$ 1 so classify as 1. 1

(c)

$$dist = \frac{g(\boldsymbol{x})}{|\boldsymbol{w}|} = \frac{10 + 2 \times 3 + \ -7 \times 2 + 1 \times 1}{\sqrt{4 + 49 + 1}}$$ 1 1 1

$$= \frac{3}{\sqrt{54}} = \frac{1}{\sqrt{6}} = \frac{\sqrt{6}}{6} = 0.408$$ 2

11

5 (a) (i) $\frac{1}{\sqrt{5}}(1,-2)$ or $\frac{1}{\sqrt{5}}(-1,2)$ = (0.447, -0.894) 2

(ii) $\frac{1}{\sqrt{5}}(2,1)$ or $\frac{1}{\sqrt{5}}(-2,-1)$ = (0.894,0.447) 3 ← also accept (0,0) or undefined since variance is zero

(b) $v_1 = -2 \times 0.667 + 3 \times 0.667 + 0.333 = 1$ 1

$$v_2 = -2 \times 0.596 + 3 \times -0.298 + 1 \times 0.745 = -1.341$$ 1

$$v_3 = -2 \times 0.333 + 3 \times 0.667 + 1 \times -0.667 = 0.667$$ 1

So final result is $1p_1 - 1.341p_2 + 0.668p_3$ 2

10

NB – relies on orthogonality, which should be there for principal components (but isn't here, by mistake), so actually this solution is incorrect, but will be accepted. In this case, correct answer from matrix equation solving. Either outcome was awarded full marks, and partial marks for working as appropriate.

Students solving via matrix inversion:

$$\begin{pmatrix} 0.667 & 0.596 & 0.333 \\ 0.667 & -0.298 & 0.667 \\ 0.333 & 0.745 & -0.667 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} -2 \\ 3 \\ 1 \end{pmatrix}$$ 2

[Any method of solving this] 1

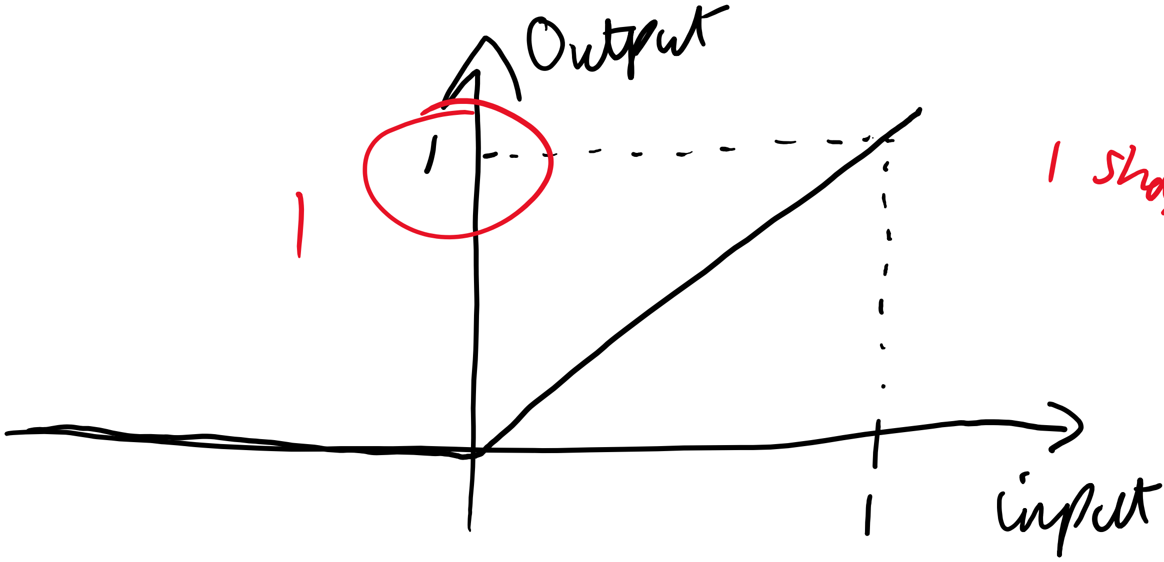
Gives

7.62p1 – 8.12p2 – 6.76p3 2

1

6 (a) A linear function applied to a linear function remains linear, and therefore no additional complexity is added by incorporating more functions. The resulting output would be a linear combination of all the inputs, which misses the value of the neural network approach. 1

(b)



1

1 shape

c) $0.2 \times 1 + 0.3 = 0.5$ so $y_1 = 0.622$ 1

$0.2 \times -0.8 - 0.1 = -0.26$ so $y_2 = 0.435$ 1

$0.622 \times 0.1 - 0.8 \times 0.435 - 0.5 = -0.786$ so $z_1 = 0.313$ 1

Residual (t1 – z1) is therefore 0.487 1

$$s = \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1}$$

$$s' = -(1+e^{-x})^{-2}(-e^{-x})$$

$$= \frac{e^{-x}}{(1+e^{-x})^2}$$ 1

Therefore

$$f'(-0.786) = 0.215$$ 1

So $\frac{\partial C}{\partial w} = \underbrace{-0.215 \times 0.435 \times 0.487}_{1} = \underbrace{-0.0455}_{1}$

accept if 2x (Q didn't specify $\frac{1}{2}$ in cost fn)

↓

−0.091

12

7. Two means at (–1, 1) and (1,-1) place dividing line on y = x. So x < y goes to mean 1, x > y goes to mean 2.

Mean 1 receives:

	x	y
	-1.5	0.9
	0.3	0.7
	-0.2	0.6
	0.2	1.4
	-0.8	0.7
Total	**-2**	**4.3**

2

So new mean 1 is at (-0.4, 0.86)

Mean 2 receives:

	x	y
	0.5	-0.9
	-0.1	-1.2
	1.3	-1.1
	1.2	-1.6
	0.6	-0.3
Total	**3.5**	**-5.1**

2

So new mean 2 is at (0.7,-1.02)

8

8

Drop the exp() function outside – this is monotonic, so has no effect on minimisation.

$$C = x^2 + 0.7y^2 + 0.9xy$$

$$\frac{\partial C}{\partial x} = 2x + 0.9y$$

$$\frac{\partial C}{\partial y} = 1.4y + 0.9x$$

Start at (0.8, 0.8)

Vary x so $\frac{\partial C}{\partial x} = 0$:

$$0 = 2x + 0.9 \times 0.8$$

$$x = \frac{-0.9 \times 0.8}{2} = -0.36$$

Vary y so $\frac{\partial C}{\partial y} = 0$:

$$y = \frac{0.9 \times 0.36}{1.4} = 0.231$$

So the steps are (0.8, 0.8) -> (-0.36, 0.8) -> (-0.36, 0.231)

6

Support vector 1 in higher dim space is $\boldsymbol{a} = \begin{pmatrix} 0 \\ 0.8 \\ 0 \end{pmatrix}$

Support vector 2 in higher dim space is $\boldsymbol{b} = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$

Classification point in higher dim space is $\boldsymbol{c} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$. Is this more or less than 50% of the way from SV1 to SV2?

$$\boldsymbol{d} = \boldsymbol{b} - \boldsymbol{a} = \begin{pmatrix} -1 \\ -1.8 \\ 1 \end{pmatrix}$$

$$\boldsymbol{e} = \boldsymbol{c} - \boldsymbol{a} = \begin{pmatrix} -1 \\ -0.8 \\ 0 \end{pmatrix}$$

Project e into direction of d and see what fraction of d this is:

$$2 \quad \frac{\boldsymbol{e}.\hat{\boldsymbol{d}}}{|\boldsymbol{d}|} = \frac{\boldsymbol{e}.\boldsymbol{d}}{\boldsymbol{d}.\boldsymbol{d}} = \frac{1 + 1.8 \times 0.8 + 0}{1 + 1.8^2 + 1} = \frac{2.44}{5.24} = 0.465$$

9

< 0.5, so closer to $\boldsymbol{a}$. Therefore classified as 0.

Or:

9 alt:

$$\begin{array}{ccc} 0 & 1 & ? \\ \begin{pmatrix} 0 \\ 0.8 \\ 0 \end{pmatrix} & \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} & \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \end{array}$$

$$g(\underline{x}) = \underline{w}^T \underline{x} + w_0$$

$$\underline{w} = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ 0.8 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ -1.8 \\ 1 \end{pmatrix}$$ 1

$$\underline{x}_m = \frac{1}{2}\left[\begin{pmatrix} 0 \\ 0.8 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}\right] = \begin{pmatrix} -0.5 \\ -0.1 \\ 0.5 \end{pmatrix} \qquad g(\underline{x}_m) = 0$$ 1

$$w_0 = -\underline{w}^T \underline{x}_m = -\begin{pmatrix} -1 \\ -1.8 \\ 1 \end{pmatrix}^T \cdot \begin{pmatrix} -0.5 \\ -0.1 \\ 0.5 \end{pmatrix}$$

$$= -(0.5 + 0.18 + 0.5)$$

$$= -1.18$$ 1

$$g(\underline{x}) = \begin{pmatrix} -1 \\ -1.8 \\ 1 \end{pmatrix} \cdot \underline{x} - 1.18$$

1 Try $\begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}$: $g\begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} = 1 - 1.18 = -0.18 < 0$ 1

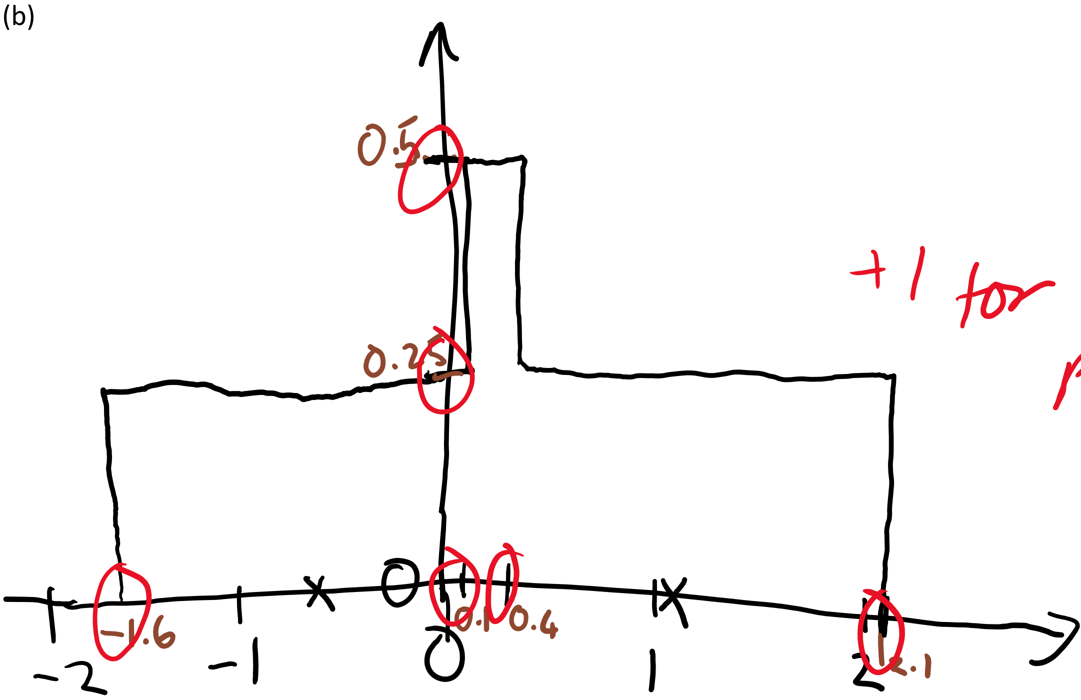
so classify as 0 1

9

10 (a) (i) A small window captures the fine details and is accurate in space 1

(ii) A large window averages more samples so gives more accurate probability across the range 1

(b)



+1 for each point

1 if all points are wrong but shape OK.

(c)

Need all points $0.75 < x_1 < 1.25$ and $1.75 < x_2 < 2.25$. There are 6 of these – (0.8, 2.1), (1.2, 2.2), (1.1, 2.0), (1.1, 2.1), (0.8, 2.2) and (0.8, 1.8).

Therefore, probability is $p = \underbrace{\frac{6}{200\times0.25} = \frac{6}{50}}_{1} = \underbrace{0.12.}_{2}$

16