Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2017-18-a

QUESTION 1

Assignment Number: 1
Student Name: Talla Aravind Reddy
Roll Number: 14746
Date: September 10, 2017

Part 1

Let $z = (x_1, y_1)$. We are given $z_r = (0, 1)$ and $z_g = (1, 0)$.
For the decision boundary, $d(z, z_r) = d(z, z_g)$

$\implies \langle z - z_r, U(z - z_r)\rangle = \langle z - z_g, U(z - z_g)\rangle$ where $U = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$

$\implies 3x_1^2 + (y_1 - 1)^2 = 3(x_1 - 1)^2 + y_1^2$

$\implies 3x_1^2 - 3(x_1 - 1)^2 = y_1^2 - (y_1 - 1)^2$

$\implies 6x_1 - 3 = 2y_1 - 1$

$\implies 3x_1 = y_1 + 1$

$\implies 3x = y + 1$ is the decision boundary.

Part 2

For the decision boundary, $d(z, z_r) = d(z, z_g)$

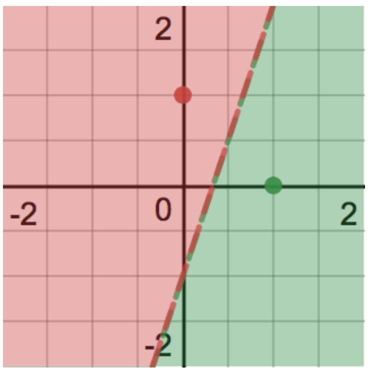
$\implies \langle z - z_r, U(z - z_r)\rangle = \langle z - z_g, U(z - z_g)\rangle$ where $U = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$

$\implies x_1^2 = (x_1 - 1)^2$

$\implies 2x_1 - 1 = 0$

$\implies x_1 = \frac{1}{2}$

$\implies x = \frac{1}{2}$ is the decision boundary.



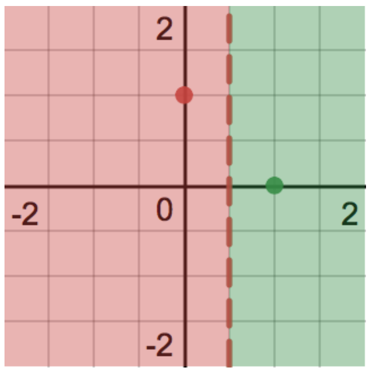


Figure 1: The figure on the left shows the decision boundary for part 1 i.e $3x = y + 1$.
The figure on the right shows the decision boundary for part 2 i.e $x = \frac{1}{2}$.

Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2017-18-a

QUESTION 2

Assignment Number: 1
Student Name: Talla Aravind Reddy
Roll Number: 14746
Date: September 10, 2017

One likelihood distribution which leads to $\hat{\mathbf{w}}_{cls}$ as the MAP estimate for the model is the gaussian distribution with mean as $\langle \mathbf{w}, \mathbf{x}^i \rangle$

$$\mathbb{P}[y|\mathbf{x}^i, w] = \mathcal{N}(\langle \mathbf{w}, \mathbf{x}^i \rangle, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \langle \mathbf{w}, \mathbf{x}^i \rangle)^2}{2\sigma^2} \right)$$

The prior distribution encapsulates the fact that $||\mathbf{w}||_2 \leq r$. This can be achieved by a uniform distribution over the d-dimensional ball with origin as centre and radius r.

$$\mathbb{P}[\mathbf{w}] = \mathcal{U}(\mathbf{0}, r) = \begin{cases} \dfrac{\Gamma(\frac{d}{2} + 1)}{\pi^{\frac{d}{2}} r^d} & \text{if } ||\mathbf{w}||_2 \leq r \\ 0 & \text{if } ||\mathbf{w}||_2 > r \end{cases}$$

QUESTION

3

Assignment Number: 1
Student Name: Talla Aravind Reddy
Roll Number: 14746
Date: September 10, 2017

The likelihood part of the estimate is same as the previous question, therefore one likelihood distribution is the gaussian distribution with mean as $\langle \mathbf{w}, \mathbf{x}^i \rangle$

$$\mathbb{P}[y|\mathbf{x}^i, w] = \mathcal{N}(\langle \mathbf{w}, \mathbf{x}^i \rangle, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\bigg(- \frac{(y - \langle \mathbf{w}, \mathbf{x}^i \rangle)^2}{2\sigma^2} \bigg)$$

For the prior, we can take a multivariate gaussian distribution with mean at $\mathbf{0}$

$$\mathbb{P}[\mathbf{w}] = \mathcal{N}(\mathbf{w}; \mathbf{0}, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\bigg(- \frac{1}{2} \mathbf{w}^T \Sigma^{-1} \mathbf{w} \bigg)$$

Here $\Sigma = \sigma^2 \begin{bmatrix} \frac{1}{\alpha_1} & 0 & \dots & \dots & 0 \\ 0 & \frac{1}{\alpha_2} & 0 & \dots & 0 \\ \vdots & 0 & \ddots & 0\dots & 0 \\ \vdots & \vdots & \vdots & \frac{1}{\alpha_{d-1}} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\alpha_d} \end{bmatrix}$ i.e a diagonal matrix with it's i'th diagonal entry as $\frac{\sigma^2}{\alpha_i}$ where σ^2 is the variance of the likelihood distribution.

$$\log \mathbb{P}[\mathbf{w}|\mathbf{X}, \mathbf{y}] = C - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y^i - \langle \mathbf{w}, \mathbf{x}^i \rangle)^2 - \frac{1}{2\sigma^2} \sum_{j=1}^{d} \alpha_j \mathbf{w}_j^2$$

$$\implies \hat{\mathbf{w}}_{\text{MAP}} = \arg\min \sum_{i=1}^{n} (y^i - \langle \mathbf{w}, \mathbf{x}^i \rangle)^2 + \sum_{j=1}^{d} \alpha_j \mathbf{w}_j^2$$

Closed form expression:

We need to minimise

$$\mathcal{L} = \sum_{i=1}^{n} (y^i - \langle \mathbf{w}, \mathbf{x}^i \rangle)^2 + \sum_{j=1}^{d} \alpha_j \mathbf{w}_j^2$$

$$\mathcal{L} = ||\mathbf{X}\mathbf{w} - \mathbf{Y}||^2 + \sum_{j=1}^{d} \alpha_j \mathbf{w}_j^2$$

$\nabla_w \mathcal{L} = 2\mathbf{X}^T(\mathbf{X}\mathbf{w} - \mathbf{Y}) + 2\mathbf{D}_\alpha \mathbf{w}$ where D_α is the diagonal matrix with entries $\alpha_1, \alpha_2 \dots \alpha_d$

$$\nabla_w \mathcal{L} = 2((\mathbf{X^T X} + \mathbf{D}_\alpha)\mathbf{w} - \mathbf{X^T Y})$$

$$\nabla_w \mathcal{L} = 0$$

$$\iff (\mathbf{X^T X} + \mathbf{D}_\alpha)\mathbf{w} - \mathbf{X^T Y} = 0$$

$$\iff (\mathbf{X^T X} + \mathbf{D}_\alpha)\mathbf{w} = \mathbf{X^T Y}$$

$$\iff w = (\mathbf{X^T X} + \mathbf{D}_\alpha)^{\mathbf{-1}} \mathbf{X^T Y}$$

Indian Institute of Technology Kanpur
CS771 Introduction to Machine Learning, 2017-18-a

Assignment Number: 1
Student Name: Talla Aravind Reddy
Roll Number: 14746
Date: September 10, 2017

QUESTION 4

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis feugiat vehicula dolor, sed ultricies leo. Phasellus euismod dictum felis in euismod. Proin pretium vel neque in placerat. Proin imperdiet egestas vulputate. Etiam faucibus accumsan ante non viverra. Duis ultrices ac odio vel sodales. In maximus gravida dolor, ut commodo lacus. Pellentesque ante massa, venenatis id aliquam et, posuere sed dui. Duis dignissim justo sit amet augue posuere fringilla. Suspendisse at nisi gravida, mattis justo sit amet, elementum elit. Praesent et massa ornare, consequat dui eget, ornare risus. Duis est nibh, sollicitudin nec mattis non, mattis in leo. Donec finibus justo sed massa sagittis, non fermentum nibh dictum. Pellentesque et congue purus. Donec porta pretium porttitor.

Morbi euismod risus eu tortor ornare malesuada. Nunc sed sollicitudin neque, efficitur rhoncus tellus. Cras malesuada augue arcu. Sed sem odio, tincidunt quis laoreet ac, facilisis ut nibh. Quisque gravida dolor at egestas aliquam. Aenean mollis massa sit amet enim mattis, vel fermentum tortor facilisis. Donec pellentesque est velit, vitae posuere lorem tristique ut.