1.a. – conditional independence given class

1b – discriminative

1c – 25%

1d – $P(A|B) = \dfrac{P(A,B)}{\sum_{a \in A} P(A,B)}$

1e – 1

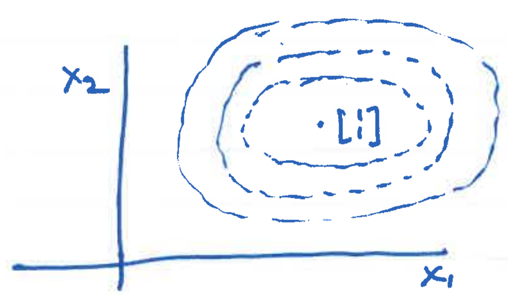
1f – True

1.g $P(A|B) = \dfrac{P(B/A)\,P(A)}{P(B)}$

$$= \frac{0.07 \times 0.10}{0.05} = 0.07 \times 2 = 0.14$$

2a $\mu_{d \times 1}$, $\Sigma_{d \times d}$

2b



ellipses aligned along x_1 (major axis) and x_2 (minor axis)

2c (COMP 3432) $\displaystyle\sum_{i=1}^{D} \frac{(x_i-\mu_i)^3}{\sigma_i^2}$

2c (COMP 4432)

2 B Gaussian density is

$$\frac{1}{(2\pi)\,|\Sigma_d|^{1/2}} \exp\left(-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix} - \begin{bmatrix}\mu_1\\\mu_2\end{bmatrix}\right)^T \begin{bmatrix}\sigma_1^2 & 0\\0 & \sigma_2^2\end{bmatrix}^{-1} \begin{bmatrix}x_1-\mu_1\\x_2-\mu_2\end{bmatrix}\right)$$

$$= \frac{1}{(2\pi)\,(\sigma_1^2\sigma_2^2)^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_1-\mu_2)^2}{\sigma_2^2}\right)\right]$$

$$= \frac{1}{(2\pi)^{1/2}\sigma_1} \exp\left(-\frac{1}{2}\frac{(x_1-\mu_1)^2}{\sigma_1^2}\right) \frac{1}{(2\pi)^{1/2}\sigma_2} \exp\left(-\frac{1}{2}\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)$$

$$= \mathcal{N}(x_1;\mu_1,\sigma_1^2)\,\mathcal{N}(x_2;\mu_2,\sigma_2^2)$$

(3) log likelihood, Find σ s.t $P(D)$ is maximum

or $\log P(D) = \sum_{i=1}^{N} \log \frac{1}{2\sigma} \exp\left(-\frac{|x_i|}{\sigma}\right)$

$\uparrow$ ($\because$ i.i.d condition)

$$= \sum_{i=1}^{N} \log \frac{1}{2\sigma} - \sum_{i=1}^{N} \frac{|x_i|}{\sigma}$$

$$= -\sum_{i=1}^{N} \log 2\sigma - \sum_{i=1}^{N} \frac{|x_i|}{\sigma}$$

or minimize

$$F(\sigma) = \sum_{i}^{N} \log 2\sigma + \sum_{i}^{N} \frac{|x_i|}{\sigma}$$

$$\frac{\partial F(\sigma)}{d\sigma} = \sum_{i}^{N} \frac{\cancel{2}}{\cancel{2}\sigma} + \sum_{i}^{N} \frac{|x_i|(-1)}{\sigma^2}$$

$$\frac{1}{\sigma}\left[N - \sum_{i}^{N} \frac{|x_i|}{\sigma}\right] = 0$$

$$\sigma = \sum_{i}^{N} \frac{|x_i|}{N}$$

(4) - l_1 - regularization

- l_1 solution is sparse (most of the coeffient are zero)
- bigger λ - more and more coefficient will become zero

. This will allow us to do feature selection when we have lots of features

(5) As per given Hint

Decision boundary is all the x s.t

$$P(y=1|x) = P(y=0|x)$$

$$\Leftrightarrow \sigma(w^T x) = 1 - \sigma(w^T x)$$

$$\Leftrightarrow \frac{1}{1+e^{-w^T x}} = \frac{e^{-w^T x}}{1+e^{-w^T x}}$$

$$\Leftrightarrow 1 = e^{-w^T x}$$

$$\Leftrightarrow \log 1 = \log e^{-w^T x} \quad (\because \text{taking log on both side})$$

$$\Leftrightarrow 0 = -w^T x$$

$$\Leftrightarrow w^T x = 0$$ which is a equation of Line in D dimension passing through origin