Q1 6 Points $p(y=c|x) = p(x|y=c)\, p(y=c) / p(x)$

If $p(x|y=c) = \mathcal{N}(x|\mu_c, \Sigma_c)$ and $\Sigma_c = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_D^2 \end{bmatrix}_{D\times D}$

then Factor $\mathcal{N}(x|\mu_c, \Sigma_c)$ to show that

$p(x|y=c)$ can be factored into 1-D Gaussian

It should follow from property of exp,

Σ_c being diagonal $|\Sigma_c| = \det = \sigma_1^2 \sigma_2^2 \cdots \sigma_D^2$

Q2 -

2 point

4.1 [redacted]

$E[X] = 0$, $Var[X] = (1-(-1))^2/12 = 1/3$

$E[Y] = E[X^2] = var[X] + (E[X])^2 = 1/3 + 0 = 1/3$

$E[XY] = \int_{-1}^{1} x y\, p(x)\, dx = \int_{-1}^{1} \frac{x^3}{2} dx = \frac{1}{2} \int_{-1}^{1} \frac{x^4}{4} dx = 0$

$\rho = \frac{cov[x,y]}{\sigma_x \sigma_y} = \frac{E[XY] - E[X]E[Y]}{\sigma_x \sigma_y} = 0$

4.14 (1 point each) 4 points total

(a) prior on mean μ

$$P(\mu) = (2\pi s^2)^{-1/2} \exp\left[-\frac{1}{2}\frac{(\mu-m)^2}{s^2}\right]$$

$D = \{x_i\}_{i=1}^{n}$ are taken I-I-D

then $P(D|\mu) = \prod_{i=1}^{n} N(\mu, \sigma^2)$

$$= \underbrace{\prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2}}_{A} \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

$$= A \prod_{i=1}^{n} \exp\left(-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right)$$

Hence posterior on mean $P(\mu|D) \propto P(D|\mu)\,P(\mu)$

For maximizing posterior we can take log

Hence $\log[P(D|\mu)P(\mu)] = \log A - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2}\log(2\pi s^2) - \frac{(\mu-m)^2}{2s^2}$

For MAP Estimate ~~we need~~ of μ we need to take derivative of above posterior of μ with respect to μ

$$\frac{\partial \log[P(D|\mu)P(\mu)]}{\partial \mu} = 0 - \frac{2}{2\sigma^2}(x_i-\mu)(-1) - 0 - \frac{2}{2s^2}(\mu-m)$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) - \frac{(\mu-m)}{s^2}$$

$$= -\frac{\mu}{s^2} + \frac{m}{s^2} + \frac{1}{\sigma^2}\sum_{i=1}^{n}x_i - \frac{n\mu}{\sigma^2}$$ [let $\bar{x} = \sum_{i=1}^{n} x_i$

let set it zero and solve for μ

$$0 = -\frac{\mu}{s^2} + \frac{m}{s^2} + \frac{n\bar{x}}{\sigma^2} - \frac{n}{\sigma^2}\mu$$

$$\mu\left(\frac{n}{\sigma^2}+\frac{1}{s^2}\right)=\frac{1}{\sigma^2}\bar{x}n+\frac{m}{s^2}$$

Hence $$\hat{\mu}_{MAP}=\frac{\frac{1}{\sigma^2}n\bar{x}+\frac{m}{s^2}}{\frac{n}{\sigma^2}+\frac{1}{s^2}}=\frac{ns^2\bar{x}+\sigma^2 m}{ns^2+\sigma^2}$$

$$\boxed{=\left(\frac{ns^2}{ns^2+\sigma^2}\right)\bar{x}+\left(\frac{\sigma^2}{ns^2+\sigma^2}\right)m}$$

even mentioning this result is ok.

(b) as n increases $\frac{\sigma^2}{ns^2+\sigma^2}\to 0$ and $\frac{ns^2}{ns^2+\sigma^2}\to 1$ (L'Hospital rule)

Hence $\hat{\mu}_{MAP}=\bar{x}=\hat{\mu}_{MLE}$

i.e as we have more sample prior becomes less relevant

(c) If prior variance s^2 increases then $\frac{\sigma^2}{ns^2+\sigma^2}\to 0$, $\frac{ns^2}{ns^2+\sigma^2}\to 1$

again $\hat{\mu}_{MAP}=\hat{\mu}_{MLE}$.

When we have uninformative prior (uniform almost) we end up relying more on data.

(d) if $s^2\to 0$ $\frac{\sigma^2}{ns^2+\sigma^2}\to 1$ and $\frac{ns^2}{ns^2+\sigma^2}\to 0$

then $\hat{\mu}_{MAP}=m$

i.e if prior is concentrated (small s^2) then we already have answer and can ignore data (evidence)

$4 = 2+2$ points

4.21 (a)

decision region $R_1 = \{x : p(x|\mu_1, \sigma_1) \geq p(x|\mu_2, \sigma_2)\}$

i.e x has to satisfy

$$\frac{1}{\cancel{(2\pi)^{1/2}}\sigma_1} \exp\left(\frac{-1}{2\sigma_1^2}(x-\mu_1)^2\right) \geq \frac{1}{\cancel{(2\pi)^{1/2}}\sigma_2} \exp\left(\frac{-1}{2\sigma_2^2}(x-\mu)^2\right)$$

taking log

$$-\log\sigma_1 - \frac{1}{2\sigma_1^2}(x-\mu_1)^2 \geq -\log\sigma_2 - \frac{1}{2\sigma_2^2}(x-\mu_2)^2$$

given $\mu_1 = 0$, $\mu_2 = 1$, $\sigma_1^2 = 1$, $\sigma_2^2 = 10^6$

$$\frac{-x^2}{2} \geq -\log\sigma_2 - \frac{x^2}{2\sigma_2^2} - \frac{\mu_2^2}{2\sigma_2^2} + \frac{\mu_2 x}{\sigma_2^2}$$

$$-x^2 \geq -2\log\sigma_2 - \frac{x^2}{\sigma_2^2} + \frac{2x}{\sigma_2^2} - \frac{1}{\sigma_2^2} \quad -(1)$$

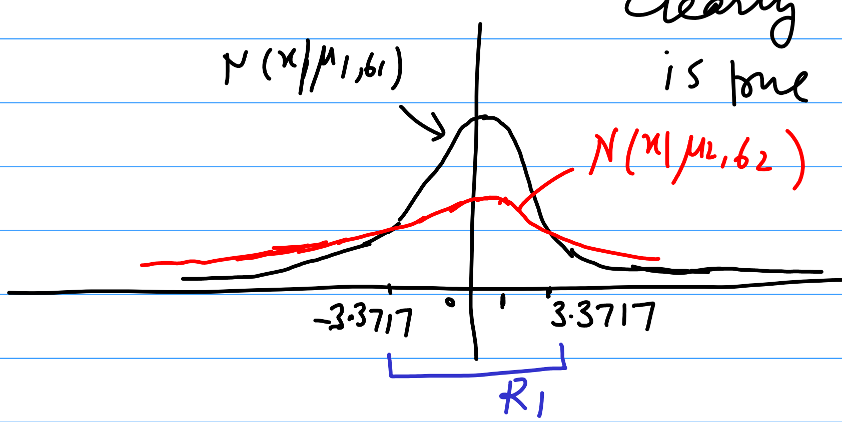
$$\left(\frac{1}{\sigma_2^2} - 1\right)x^2 - \frac{2}{\sigma_2^2}x + \frac{1}{\sigma^2} + 2\log\sigma_2 \geq 0$$

using any quadratic solver for equality part we get

$x = \pm 3.3717$

$(x+3.3717)(-x+3.3717) \geq 0$

Clearly above inequality is true for $x \in [-3.3717\ +3.3717]$



(b) If $\sigma_2 = 1$ then equation (1) becomes

$$-x^2 \geq 0 - x^2 + 2x - 1$$

$$-2x + 1 \geq 0$$

$$2x - 1 \leq 0$$

$$x \leq 1/2$$

Hence $x \in R_1$ if $x \leq .5$

(2 = 1+1) point

4.22

$$P(y=c|X) = \frac{P(X|y=c)\,P(y=c)}{P(x)}$$

(a) using given values, we should get

$$P(y=1/X_1) = 0.46$$

$$P(y=2|X_1) = 0.145)$$

$$P(y=3|X_1) = 0.39$$

also

$$R(\alpha_1|X_1) = 0.145 + 0.39 = 0.53$$

$$R(\alpha_2|X_2) = 0.85 \text{ and } R(\alpha_3|X_3) = 0.60$$

Clearly class 1 has minimal risk/maximum posterior probability.

(b) Similarly

$$P(y=1|X_2) = 0.45, \quad P(y=2|X_2) = 0.46$$

$$P(y=3|X_2) = 0.09$$

Also $R(\alpha_1|X_2) = 0.55$, $R(\alpha_2|X_2) = .54$, $R(\alpha_3|X_2) = 0.91$

clearly class 2