HW #2 solutions

3.3 Define $E[X] = \mu$ and $V[X] = \sigma^2$, then for $Y = aX + b$ we have:

$$\begin{cases} E[Y] = aE[X] + b = a\mu + b, \\ V[Y] = a^2 V[X] = a^2\sigma^2. \end{cases}$$

Next, $E[XY] = E[aX^2 + bX] = a(\mu^2 + \sigma^2) + b\mu$.

$$Cov[X, Y] = E[XY] - E[X]E[Y] =$$

$$= a(\mu^2 + \sigma^2) + b\mu - \mu(a\mu + b) = \underline{\underline{a\sigma^2}}.$$

Finally, $$\rho = \frac{Cov[X,Y]}{\sqrt{V[X]V[Y]}} = \frac{a\sigma^2}{\sqrt{\sigma^2}\sqrt{a^2\sigma^2}} =$$

$$= \frac{a}{|a|}.$$

Thus, $\rho = 1$ if $a > 0$ and $\rho = -1$ if $a < 0$.

$a = |a|$ $\qquad$ $a = -|a|$

3.6 Since Σ is square and symmetric, we can write

$$\Sigma = \mathcal{U} \underbrace{\Lambda}_{= \operatorname{diag}(\lambda_1, \ldots, \lambda_d)} \mathcal{U}^T .$$

Then $\Sigma^{-1} = (\mathcal{U}^T)^{-1} \Lambda^{-1} \mathcal{U}^{-1} = \mathcal{U} \Lambda^{-1} \mathcal{U}^T =$

$$= \sum_{i=1}^{d} \frac{1}{\lambda_i} \vec{u}_i \cdot \vec{u}_i^T .$$

Next, $(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu}) =$

$$= \sum_{i=1}^{d} \frac{1}{\lambda_i} (\vec{x} - \vec{\mu}) \cdot \vec{u}_i \times \vec{u}_i \cdot (\vec{x} - \vec{\mu}) \ (=)$$

Define $y_i = \vec{u}_i \cdot (\vec{x} - \vec{\mu})$, or

$$\underbrace{\vec{y}}_{\text{d-dim vector}} = \mathcal{U}(\vec{x} - \vec{\mu}) \Rightarrow \vec{x} = \mathcal{U}^T \vec{y} + \vec{\mu} .$$

$$(=) \sum_{i=1}^{d} \frac{y_i^2}{\lambda_i} .$$

The Jacobian is given by

$$J_{ij} = \frac{\partial x_i}{\partial y_j} = (\mathcal{U}^T)_{ij} = \mathcal{U}_{ji} .$$

Thus, $J = \mathcal{U}^T \Rightarrow \det(J) = 1$.

Finally, we obtain:

$$\int d\vec{x}\, e^{-\frac{1}{2}(\vec{x}-\vec{\mu})\Sigma^{-1}(\vec{x}-\vec{\mu})} =$$

$$= \int dy_1 \ldots dy_d \underbrace{|J|}_{=1} e^{-\frac{1}{2}\sum_{i=1}^{d} \frac{y_i^2}{\lambda_i}} =$$

$$= \prod_{i=1}^{d} \sqrt{2\pi\lambda_i} = (2\pi)^{d/2} \underbrace{|\Sigma|}_{\det(\Sigma)}^{1/2}$$

4.2

(a)
$$\begin{cases} \underset{\text{likelihood}}{p(\mathcal{D}|\mu,\sigma^2)} = e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n, \\ \underset{\text{prior}}{p(\mu)} = \frac{1}{\sqrt{2\pi s^2}}\, e^{-\frac{(\mu-m)^2}{2s^2}} \end{cases}$$

Then $\overbrace{\log p(\mu) + \log p(\mathcal{D}|\mu,\sigma^2)}^{=:\,\mathcal{J}} =$

$$= -\frac{1}{2}\log(2\pi s^2) - \frac{(\mu-m)^2}{2s^2} -$$

$$- \frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2.$$

$$\frac{\partial \mathcal{J}}{\partial \mu} = -\frac{\mu-m}{s^2} + \frac{1}{\sigma^2}\sum_i (x_i-\mu) =$$

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

$$\left.\frac{\partial \mathcal{J}}{\partial \mu}\right|_{\mu_{MAP}} = 0 = \frac{m-\mu_{MAP}}{s^2} + \frac{n(\bar{x}-\mu_{MAP})}{\sigma^2},$$

which yields

$$\boxed{\mu_{MAP} = \frac{ns^2}{ns^2+\sigma^2}\,\bar{x} + \frac{\sigma^2}{ns^2+\sigma^2}\,m}.$$

(b) Clearly, as $n \to \infty$

$$\mu_{MAP} \to \underbrace{\bar{x}}_{MLE}$$

(c) As $s^2 \uparrow$, $\frac{ns^2}{ns^2+\sigma^2} \to 1$ even for moderate n,

and $\frac{\sigma^2}{ns^2+\sigma^2} \to 0$.

Thus, $\mu_{MAP} \to \bar{x}$ again.

Uninformative ('flat') prior does not influence the outcome.

(d) As $s^2 \downarrow$, $\frac{ns^2}{ns^2+\sigma^2} \to 0$ and

$\frac{\sigma^2}{ns^2+\sigma^2} \to 1$, yielding

$$\mu_{MAP} \to m$$

We already 'know' the answer and ignore the data.

4.6 <u>Prior</u>: $p(\theta) = \sum_{k=1}^{\overset{\text{\# components}}{K}} \underbrace{p(\theta | z=k)}_{\text{or simply } p(\theta|k)} p(k)$, where z ← hidden var

$p(k) = \pi_k$ are the prior mixing weights.

Now, consider the posterior:

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{p(\mathcal{D})} = \frac{p(\mathcal{D}|\theta)\sum_k p(k)p(\theta|k)}{\int d\theta' p(\mathcal{D}|\theta')\sum_{k'} p(k')p(\theta'|k')} =$$

$$= \frac{\sum_k p(k)p(\mathcal{D},\theta|k)}{\sum_{k'} p(k') \underbrace{\int d\theta' p(\mathcal{D},\theta'|k')}_{p(\mathcal{D}|k')}} = \frac{\sum_k p(k)\, p(\theta|\mathcal{D},k)\, p(\mathcal{D}|k)}{\sum_{k'} p(k')\, p(\mathcal{D}|k')} =$$

$$= \sum_k \underbrace{\frac{p(k)p(\mathcal{D}|k)}{\sum_{k'} p(k')p(\mathcal{D}|k')}}_{\substack{p(k|\mathcal{D}) = \tilde{\pi}_k, \\ \text{posterior mixing} \\ \text{weights}}} \underbrace{p(\theta|\mathcal{D},k)}_{\substack{\text{posterior for } \theta \\ \text{given } k}}.$$

In summary,

$$p(\theta|\mathcal{D}) = \sum_{k=1}^{K} p(\theta|\mathcal{D},k)\, p(k|\mathcal{D}).$$

This is of the same form as the prior, especially if $p(\theta|k)$ is conjugate to $p(\theta|\mathcal{D},k)$.