Chapter 11. Mixture models and the EM algorithm

11.1.

$$\mathcal{T}(x|\mu,\sigma,\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}(1+\frac{1}{\nu}(\frac{x-\mu}{\sigma})^2)^{-\frac{\nu+1}{2}}$$

$$\mathcal{N}(x|\mu,\frac{\sigma^2}{z}) = \frac{\sqrt{z}}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2 z}{2\sigma^2}}$$

$$\mathrm{Ga}(z|\frac{\nu}{2},\frac{\nu}{2}) = \frac{\frac{\nu}{2}^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}z^{\frac{\nu}{2}-1}e^{-z\frac{\nu}{2}}$$

$$\int_0^\infty \mathcal{N}(x|\mu,\frac{\sigma^2}{z})\mathrm{Ga}(z|\frac{\nu}{2},\frac{\nu}{2})dz$$

$$= \int_0^\infty \frac{\nu^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}}\sqrt{2\pi}\sigma}e^{-z(\frac{\nu}{2}+\frac{(x-\mu)^2}{2\sigma^2})}z^{\frac{\nu-1}{2}}dz$$

$$= \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}\sigma}\frac{1}{(1+\frac{1}{\nu}(\frac{x-\mu}{\sigma})^2)^{\frac{\nu+1}{2}}} = \mathcal{T}(x|\mu,\sigma,\nu).$$

11.2.

$$\frac{\partial}{\partial\boldsymbol{\mu}_k}l(\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) = -\frac{1}{2}\sum_i r_{ik}[\frac{\partial}{\partial\boldsymbol{\mu}_k}(\mathbf{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_k)]$$

$$= \frac{1}{2}\sum_i r_{ik}[(\boldsymbol{\Sigma}_k^{-1}+\boldsymbol{\Sigma}_k^{-T})(\mathbf{x}_i-\boldsymbol{\mu}_k)] = 0$$

$$\Rightarrow \boldsymbol{\mu}_k = \frac{\sum_i r_{ik}\mathbf{x}_i}{r_k}.$$

$$\frac{\partial}{\partial \boldsymbol{\Sigma}_k} l(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = -\frac{1}{2}\sum_i r_{ik}\boldsymbol{\Sigma}_k^{-T} - \frac{1}{2}\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T = 0$$

$$\Rightarrow \boldsymbol{\Sigma}_k = \frac{\sum_i r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)(\mathbf{x}_i - \boldsymbol{\mu}_k)^T}{r_k} = \frac{\sum_i r_{ik}\mathbf{x}_i\mathbf{x}_i^T}{r_k} - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^T.$$

11.3.

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{t-1}) = \sum_i\sum_k r_{ik}[\log \pi_k + \sum_j (x_j \log \mu_{kj} + (1-x_j)\log(1-\mu_{kj}))]$$

$$\frac{\partial Q}{\partial \boldsymbol{\mu}} = \mathbf{0} \Rightarrow \sum_i r_{ik}(\frac{x_{ij}}{\mu_{kj}} - \frac{1-x_{ij}}{1-\mu_{kj}}) = 0 \Rightarrow \mu_{kj} = \frac{\sum_i r_{ik}x_{ij}}{\sum_i r_{ik}}.$$

With Beta(α, β) prior,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{t-1}) = \sum_i\sum_k r_{ik}[\log \pi_k + \sum_j (x_j \log \mu_{kj} + (1-x_j)\log(1-\mu_{kj}))]$$

$$+\sum_k\sum_j[(\alpha-1)\log \mu_{kj} + (\beta-1)\log(1-\mu_{kj})] = 0$$

$$\Rightarrow \sum_i r_{ik}(\frac{x_{ij}}{\mu_{kj}} - \frac{1-x_{ij}}{1-\mu_{kj}}) + \frac{\alpha-1}{\mu_{kj}} - \frac{\beta-1}{1-\mu_{kj}} = 0$$

$$\Rightarrow \mu_{kj} = \frac{(\sum_i r_{ik}x_{ij}) + \alpha - 1}{(\sum_i r_{ik}) + \alpha + \beta - 2}.$$

11.4.

$$p(\mathbf{x}_i, \boldsymbol{\theta}) = \sum_k \mathcal{T}(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \nu_k)$$

$$= \sum_k \pi_k \int \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_k, \frac{\boldsymbol{\Sigma}_k}{z_i})\mathrm{Ga}(z_i|\frac{\nu_k}{2}, \frac{\nu_k}{2})dz_i.$$

For E step,

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{t-1}) = \sum_i\sum_k r_{ik}\log \pi_k + \sum_i\sum_k r_{ik}\log p(\mathbf{x}_i|\boldsymbol{\theta}_k)$$

$$= \sum_i\sum_k r_{ik}\log \pi_k + \sum_i\sum_k r_{ik}[-\frac{D}{2}\log(2\pi) - \frac{1}{2}|\boldsymbol{\Sigma}_k| - \frac{\mathbb{E}[z_{ik}]}{2}\delta_i + \frac{\nu_k}{2}\log\frac{\nu_k}{2} - \log\Gamma(\frac{\nu_k}{2})$$

$$+\frac{\nu_k}{2}\mathbb{E}[\log z_{ik} - z_{ik}] + (\frac{D}{2} - 1)\mathbb{E}[\log z_{ik}]]$$

For M step,

$$\frac{\partial Q}{\partial \boldsymbol{\mu}} = \mathbf{0} \Rightarrow \sum_i \sum_k r_{ik}(\mathbf{x}_i - \boldsymbol{\mu}_k)\boldsymbol{\Sigma}_k \mathbb{E}[z_{ik}] = \mathbf{0}.$$

$$\Rightarrow \boldsymbol{\mu}_k = \frac{\sum_i r_{ik}\mathbf{x}_i \mathbb{E}[z_{ik}]}{\sum_i r_{ik}\mathbb{E}[z_{ik}]}.$$

$$\mathbb{E}[z_{ik}] = \frac{\nu_k + D}{\nu_k + \delta_{ik}},$$

where $\delta_{ik} = (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_k)$.

$$\Rightarrow \boldsymbol{\mu}_{k,t} = \frac{\sum_i r_{ik}\mathbf{x}_i \frac{\nu_k + D}{\nu_k + (\mathbf{x}_i - \boldsymbol{\mu}_{k,t-1})^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_{k,t-1})}}{\sum_i r_{ik} \frac{\nu_k + D}{\nu_k + (\mathbf{x}_i - \boldsymbol{\mu}_{k,t-1})^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_{k,t-1})}}$$

$$\boldsymbol{\Sigma}_{k,t} = \frac{1}{N}\sum_i [\frac{\nu_k + D}{\nu_k + (\mathbf{x}_i - \boldsymbol{\mu}_{k,t-1})^T \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_{k,t-1})} + (\mathbf{x}_i - \boldsymbol{\mu}_{k,t})(\mathbf{x}_i - \boldsymbol{\mu}_{k,t})^T].$$

11.5.

$$l(\boldsymbol{\theta}) = \sum_n \log p(\mathbf{x}_n|\boldsymbol{\theta}) = \sum_n \log(\sum_k \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k))$$

a.

$$\frac{\partial l}{\partial \boldsymbol{\mu}_k} = \sum_n \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k'} \pi_{k'}\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})}\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k) = \sum_n r_{nk}\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k).$$

b.

$$\frac{\partial l}{\partial \pi_k} = \sum_n \frac{\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{k'} \pi_{k'}\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'})} = \sum_n \frac{r_{nk}}{\pi_k}.$$

c.

$$\frac{\partial l}{\partial w_k} = \sum_j \frac{\partial l}{\partial \pi_j}\frac{\partial \pi_j}{\partial w_k} = \sum_j (\sum_n \frac{r_{nj}}{\pi_j}\frac{\partial \pi_j}{\partial w_k})$$

$$= \sum_{j \neq k}(\sum_n r_{nj})\frac{1}{\pi_j}(-\pi_j \pi_k) + \sum_n r_{nk}\frac{1}{\pi_k}\pi_k(1 - \pi_k)$$

$$= \sum_n r_{nk} - \sum_{k'} r_{nk'} \pi_k = \sum_n r_{nk} - \pi_k.$$

d.

$$\frac{\partial l}{\partial \mathbf{\Sigma}_k} = \sum_n \frac{\pi_k}{\sum_{k'} \pi_{k'} \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_{k'}, \mathbf{\Sigma}_{k'})} \frac{\partial}{\partial \mathbf{\Sigma}_k} \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \mathbf{\Sigma}_k)$$

$$= \sum_n \frac{\pi_k}{\sum_{k'} \pi_{k'} \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_{k'}, \mathbf{\Sigma}_{k'})} \frac{1}{2} \mathbf{\Sigma}_k^{-1} \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \mathbf{\Sigma}_k)[(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k) - 1]$$

$$= \frac{1}{2} \sum_n r_{nk} \mathbf{\Sigma}_k^{-1} [(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \mathbf{\Sigma}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k) - 1].$$

e.

$$\frac{\partial l}{\partial \mathbf{R}_k} = \frac{\partial l}{\partial \mathbf{\Sigma}_k} \frac{\partial \mathbf{\Sigma}_k}{\partial \mathbf{R}_k} = \frac{\partial l}{\partial \mathbf{\Sigma}_k} \cdot 2 \mathbf{R}_k^T$$

$$= \sum_n r_{nk} \mathbf{R}_k^{-1} [(\mathbf{x}_n - \boldsymbol{\mu}_k)^T \mathbf{R}_k^{-T} \mathbf{R}_k^{-1} (\mathbf{x}_n - \boldsymbol{\mu}_k) - 1].$$

11.6.

a.

$$r_{njk} = P(J_n = j, K_n = k | x_n, \boldsymbol{\theta}) = \frac{P(J_n = j, K_n = k, x_n | \boldsymbol{\theta})}{p(x_n | \boldsymbol{\theta})}$$

$$= \frac{P(J_n = j) P(K_n = k) P(x_n | \boldsymbol{\theta})}{\sum_{j'} p_{j'} [\sum_{k'} q_{k'} \mathcal{N}(x_n | \mu_{j'}, \sigma_{k'}^2)]} = \frac{p_j q_k \mathcal{N}(x_n | \mu_j, \sigma_k^2)}{\sum_{j'} p_{j'} [\sum_{k'} q_{k'} \mathcal{N}(x_n | \mu_{j'}, \sigma_{k'}^2)]}$$

b.

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{t-1}) = \mathbb{E}_{\boldsymbol{\theta}_{t-1}} [\sum_n \log p(x_n, J_n, K_n | \boldsymbol{\theta})]$$

$$= \sum_n \mathbb{E} [\log \prod_j \prod_k p(x_n, J_n, K_n | \boldsymbol{\theta})^{\mathbf{1}_{J_n = j, K_n = k}}]$$

$$= \sum_n \sum_j \sum_k r_{njk} (\log p_j + \log q_k + \log \mathcal{N}(x_n | \mu_j, \sigma_k^2)).$$

c.

$$\frac{\partial Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{t-1})}{\partial \mu_j} = \sum_n \sum_k r_{njk} (\frac{x_n - \mu_j}{\sigma_k^2}) = 0$$

$$\Rightarrow \mu_j = \frac{\sum_n \sum_k \frac{r_{njk}(x_n - \mu_j)}{\sigma_k^2}}{\sum_n \sum_k \frac{r_{njk}}{\sigma_k^2}}.$$

11.7.

a.

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{t-1}) = 1.4 \log \pi_1 + 1.6 \log(1 - \pi_1) + \log p(x_1|\boldsymbol{\theta}_1)$$
$$+0.4 \log p(x_2|\boldsymbol{\theta}_1) + 0.6 \log p(x_2|\boldsymbol{\theta}_2) + \log p(x_3|\boldsymbol{\theta}_2).$$

b.

$$\pi_1 = \frac{7}{15}, \pi_2 = \frac{8}{15}.$$

c.

$$\mu_1 = \frac{25}{7}, \mu_2 = \frac{65}{4}.$$

11.8.

a.

$$\mathbb{E}[\mathbf{x}] = \int \mathbf{x} \sum_k \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x} = \sum_k \pi_k (\int \mathbf{x} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x})$$
$$= \sum_k \pi_k \mu_k.$$

b.

$$\mathrm{Cov}[\mathbf{x}] = \int \mathbf{x}\mathbf{x}^T \sum_k \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x} = \sum_k \pi_k (\int \mathbf{x}\mathbf{x}^T \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\mathbf{x})$$

$$\mathbb{E}[\mathbf{x}\mathbf{x}^T] = \sum_k \pi_k (\mathrm{Cov}_{\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}[\mathbf{x}] + \mathbb{E}_{\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}[\mathbf{x}] \mathbb{E}_{\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}[\mathbf{x}]^T)$$
$$= \sum_k \pi_k (\boldsymbol{\Sigma}_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T).$$
$$\Rightarrow \mathrm{Cov}[\mathbf{x}] = \sum_k \pi_k (\boldsymbol{\Sigma}_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T) - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}]^T.$$

11.9.

Left 10 points and right 10 points forms two cluster respectively with their centroids.

11.10.

$$\sum_{i:z_i=k}\sum_{j:z_j=k}(x_i-x_j)^2 = \sum_{i:z_i=k}(n_k s^2 + n_k(\bar{x}_k - x_i)^2)$$

$$= n_k(n_k s^2) + n_k(n_k s^2 + n_k(\bar{x}_k - \bar{x}_k)^2) = 2n_k^2 s_k^2.$$

$$\Rightarrow J_W(\mathbf{z}) = \sum_k n_k^2 s_k^2 = \sum_k n_k^2 \frac{1}{n_k}\sum_{i:z_i=k}(x_i-\bar{x}_k)^2 = \sum_k n_k \sum_{i:z_i=k}(x_i-\bar{x}_k)^2.$$

11.11.

$$p(x,z|\boldsymbol{\theta}) = \prod_k \pi_k p(x|\boldsymbol{\theta}_k)^{\mathbf{1}_{z=k}} = \frac{h(\mathbf{x})}{Z(\boldsymbol{\theta})}e^{\boldsymbol{\theta}^T\boldsymbol{\phi}(x)}$$

where $\boldsymbol{\phi}(x) = [1, x, x^2]$, $\boldsymbol{\theta} = [-\sum_k \frac{\mu_k^2 z_k}{2\sigma_k^2}, \sum_k \frac{\mu_k z_k}{\sigma_k^2}, -\sum_k \frac{z_k}{2\sigma_k^2}]$, $h(\mathbf{x}) = 1$, $Z(\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}\prod_k \sigma_k^{z_k}}$.

11.12.

$$L_N(\mu) = \frac{1}{2\sigma^2}\sum_i z_i(y_i - \mathbf{w}^T\mathbf{x}_i)^2$$

$$\frac{\partial L_N}{\partial \mathbf{w}} = \frac{1}{2\sigma^2}\sum_i z_i(y_i - \mathbf{w}^T\mathbf{x}_i)\cdot 2\cdot(-\mathbf{x}_i^T) = \mathbf{0}$$

$$\Rightarrow \hat{\mathbf{w}} = \frac{\sum_i z_i y_i \mathbf{x}_i}{\sum_i z_i \mathbf{x}_i \mathbf{x}_i^T}.$$

11.13.

$$p(\bar{x}_j|\mu,\tau^2,\sigma_j^2) = \int \mathcal{N}(\bar{x}_j|\theta_j,\sigma_j^2)\mathcal{N}(\theta_j|\mu,\tau^2)d\theta_j = \mathcal{N}(\bar{x}_j|\mu,\tau^2+\sigma_j^2).$$

$$\Rightarrow p(\mathcal{D}|\mu,\tau^2,\boldsymbol{\sigma}^2) = \prod_j \mathcal{N}(\bar{x}_j|\mu,\tau^2+\sigma_j^2).$$

$$l(\boldsymbol{\theta}) = \sum_j(-\frac{1}{2}\log 2\pi(\tau^2+\sigma_j^2) - \frac{(\bar{x}_j-\mu)^2}{2(\tau^2+\sigma_j^2)})$$

$$\frac{\partial l}{\partial \mu} = \sum_j \frac{\bar{x}_j - \mu}{\tau^2 + \sigma_j^2} = 0 \Rightarrow \hat{\mu} = \frac{\sum_j \frac{\bar{x}_j}{\tau^2+\sigma_j^2}}{\sum_j \frac{1}{\tau^2+\sigma_j^2}}.$$

$$\frac{\partial l}{\partial \tau} = 0 \Rightarrow \sum_j \frac{\tau^2 + \sigma_j^2 - (\bar{x}_j - \mu)^2}{(\tau^2 + \sigma_j^2)^2} = 0.$$

Unfortunately, this formula is untractable.

11.14.
We have $z_i \sim \mathcal{N}(\mathbf{w}^T \mathbf{x}_i, \sigma^2)$, $y_i = \min(z_i, c_i)$. Let $d_i = \mathbf{1}_{z_i \le c_i}$.
Then the complete data log-likelihood is

$$l_c(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2} \sum_i (z_i - \mathbf{w}^T \mathbf{x}_i)^2 + C.$$

We can conclude that

$$p(z_i|\mathbf{x}_i, y_i, c_i, d_i, \mathbf{w}) = d_i \delta(z_i - y_i) + (1 - d_i) \frac{\mathcal{N}(z_i|\mathbf{w}^T \mathbf{x}_i, \sigma^2)}{1 - \Phi(\frac{c_i - \mathbf{w}^T \mathbf{x}_i}{\sigma})}.$$

The expected complete data log-likelihood is

$$\int \log p(z_i|\mathbf{w}) p(z_i|\mathbf{x}_i, y_i, c_i, d_i, \mathbf{w}_{t-1}) dz_i$$

. If $d_i = 1$, the integral equals $-\frac{1}{2\sigma^2}(y_i - \mathbf{w}^T \mathbf{x}_i)^2$.
If $d_i = 0$, we use Exercise 11.15. and for brevity denote $\mu_i = \mathbf{w}^T \mathbf{x}_i$, $\mu_{i,t-1} = \mathbf{w}_{t-1}^T \mathbf{x}_i$, $a_i = \frac{c_i - \mu_i}{\sigma}$.

$$\mathbb{E}[\log p(z_i|\mathbf{w})|z_i > c_i] = -\frac{1}{2\sigma^2}[\mu_i^2 + \mathbb{E}[z_i^2|z_i > c_i] - 2\mu_i \mathbb{E}[z_i|z_i > c_i]]$$

$$= -\frac{1}{2\sigma^2}(\mu_i^2 + \mu_{i,t-1}^2 + H(a_i) - 2\mu_i(\mu_{i,t-1}^2 + \sigma^2 + \sigma(\mu_{i,t-1} + c_i + H(a_i)))).$$

Let $\mu_{i,t-1}^2 + \sigma^2 + \sigma(\mu_{i,t-1} + c_i) H(a_i) = b_i$. Then

$$Q(\mathbf{w}, \mathbf{w}_{t-1}) = -\frac{1}{2\sigma^2} \sum_i ((y_i - \mu_i)^2 d_i + (\mu_i^2 - 2b_i \mu_i + H(a_i) + \mu_{i,t-1}^2)(1 - d_i))$$

$$= -\frac{1}{2\sigma^2} \sum_i [\mu_i^2 - 2\mu_i(d_i y_i + b_i(1 - d_i))] + C.$$

Let $e_i = y_i d_i + b_i(1 - d_i)$ and $\mathbf{e}$ be a vector consisting of e_i, then

$$Q(\mathbf{w}, \mathbf{w}_{t-1}) = -\frac{1}{2\sigma^2}(\mathbf{w}^T\mathbf{X}^T\mathbf{X}\mathbf{w} - 2\mathbf{w}^T\mathbf{X}^T\mathbf{e}) + C.$$

$$\frac{\partial Q}{\partial \mathbf{w}} = \mathbf{0} \Rightarrow \mathbf{w}^T(\mathbf{X}^T\mathbf{X}) - \mathbf{X}^T\mathbf{e} = \mathbf{0}$$
$$\Rightarrow \hat{\mathbf{w}} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{e}^T\mathbf{X}).$$

11.15.
Let $a_i = \frac{c_i - \mu_i}{\delta}$ and E be the event such that $z_i \geq c_i$. Then $E = \epsilon_i \geq a_i$.

$$\mathbb{E}[z_i|E] = \int z_i p(\epsilon_i|E)d\epsilon_i = \int z_i \frac{p(\epsilon_i, E)}{p(E)} d\epsilon_i$$
$$= \frac{1}{p(E)}\int_{a_i}^{\infty} p(\epsilon_i)(\mu_i + \sigma\epsilon_i)d\epsilon_i = \frac{1}{p(E)}\mu_i \int_{a_i}^{\infty} p(\epsilon_i)d\epsilon_i + \frac{\sigma}{p(E)}\int_{a_i}^{\infty} \epsilon_i p(\epsilon_i)d\epsilon_i.$$

Since $\epsilon_i \sim \mathcal{N}(0, 1)$ and $\frac{d}{dx}\phi(x) = -x\phi(x)$,

$$\mathbb{E}[z_i|z_i \geq c_i] = \mu_i + \frac{\sigma}{p(E)}\phi(a_i) = \mu_i \sigma H(\frac{c_i - \mu_i}{\sigma}).$$

$$\mathbb{E}[z_i^2|z_i \geq c_i] = \int z_i^2 p(\epsilon_i|E)d\epsilon_i = \frac{1}{p(E)}\int_{a_i}^{\infty}(\mu_i + \sigma\epsilon_i)^2 p(\epsilon_i)d\epsilon_i$$
$$= \mu_i^2 + \frac{2\mu_i\sigma}{p(E)}\int_{a_i}^{\infty} \epsilon_i p(\epsilon_i)d\epsilon_i + \frac{\sigma^2}{p(E)}\int_{a_i}^{\infty} \epsilon_i^2 p(\epsilon_i)d\epsilon_i$$
$$= \mu_i^2 + \frac{2\mu_i\sigma}{p(E)}\phi(a_i) + \frac{\sigma^2}{p(E)}(1 - \Phi(a_i) + a_i\phi(a_i))$$
$$= \mu_i^2 + \sigma^2 + (2\mu_i\sigma + a_i\sigma^2)H(a_i)$$
$$= \mu_i^2 + \sigma^2 + \sigma(c_i + \mu_i)H(\frac{c_i - \mu_i}{\sigma}).$$