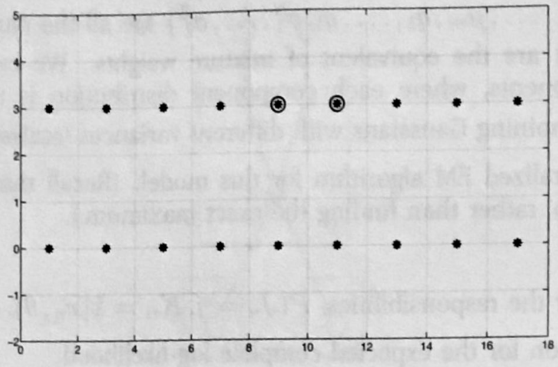

Figure 11.24 Some data points in 2d. Circles represent the initial guesses for $\mathbf{m}_1$ and $\mathbf{m}_2$.

b. Show that

$$\text{cov}[\mathbf{x}] = \sum_k \pi_k [\boldsymbol{\Sigma}_k + \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T] - \mathbb{E}[\mathbf{x}]\,\mathbb{E}[\mathbf{x}]^T \tag{11.130}$$

Hint: use the fact that $\text{cov}[\mathbf{x}] = \mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right] - \mathbb{E}[\mathbf{x}]\,\mathbb{E}[\mathbf{x}]^T$.

Exercise 11.9 K-means clustering by hand
(Source: Jaakkola.)

In Figure 11.24, we show some data points which lie on the integer grid. (Note that the x-axis has been compressed; distances should be measured using the actual grid coordinates.) Suppose we apply the K-means algorithm to this data, using $K = 2$ and with the centers initialized at the two circled data points. Draw the final clusters obtained after K-means converges (show the approximate location of the new centers and group together all the points assigned to each center). Hint: think about shortest Euclidean distance.

Exercise 11.10 Deriving the K-means cost function
Show that

$$J_W(\mathbf{z}) = \frac{1}{2}\sum_{k=1}^{K}\sum_{i:z_i=k}\sum_{i':z_{i'}=k}(x_i - x_{i'})^2 = \sum_{k=1}^{K} n_k \sum_{i:z_i=k}(x_i - \overline{x}_k)^2 \tag{11.131}$$

Hint: note that, for any μ,

$$\sum_i (x_i - \mu)^2 = \sum_i [(x_i - \overline{x}) - (\mu - \overline{x})]^2 \tag{11.132}$$

$$= \sum_i (x_i - \overline{x})^2 + \sum_i (\overline{x} - \mu)^2 - 2\sum_i (x_i - \overline{x})(\mu - \overline{x}) \tag{11.133}$$

$$= ns^2 + n(\overline{x} - \mu)^2 \tag{11.134}$$

where $s^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2$, since

$$\sum_i (x_i - \overline{x})(\mu - \overline{x}) = (\mu - \overline{x})\left((\sum_i x_i) - n\overline{x}\right) = (\mu - \overline{x})(n\overline{x} - n\overline{x}) = 0 \tag{11.135}$$

Exercise 11.11 Visible mixtures of Gaussians are in the exponential family
Show that the joint distribution $p(x, z|\boldsymbol{\theta})$ for a 1d GMM can be represented in exponential family form.