Problem 17.1

How many parameters are needed to create a 1D mixture of Gaussians with $n = 5$ components (equation 17.4)? State the possible range of values that each parameter could take.

$$P(x) = \sum_{i=1}^{n} \lambda_n \cdot \text{Norm}_x[\mu_n, \sigma_n^2] \tag{17.4}$$

Problem 17.2

A function is concave if its second derivative is less than or equal to zero everywhere. Show that this is true for the function $g(x) = \log(x)$.

Problem 17.3

For convex functions, Jensen's inequality works the other way around.

$$g(\mathbb{E}[x]) \leq \mathbb{E}[g(x)] \tag{17.31}$$

A function is convex if its second derivative is greater than or equal to zero everywhere. Show that the function $g(x) = x^{2n}$ is convex for arbitrary $n \in [1, 2, \ldots]$. Use this result with Jensen's inequality to show that the square of the mean $\mathbb{E}[x]$ of a distribution $P(x)$ must be less than or equal to its second moment $\mathbb{E}[x^2]$.

Problem 17.4

Show that the ELBO, as expressed in equation 17.18, can alternatively be derived from the KL divergence between the variational distribution $q(\mathbf{z}|\mathbf{x})$ and the true posterior distribution $P(\mathbf{z}|\mathbf{x}, \boldsymbol{\phi})$:

$$D_{KL}[q(\mathbf{z}|\mathbf{x})||P(\mathbf{z}|\mathbf{x}, \boldsymbol{\phi})] = \int q(\mathbf{z}|\mathbf{x}) \log\left[\frac{q(\mathbf{z}|\mathbf{x})}{P(\mathbf{z}|\mathbf{x}, \boldsymbol{\phi})}\right] d\mathbf{z} \tag{17.32}$$

Start by using Bayes' rule (equation 17.19).

Problem 17.5

The reparameterization trick computes the derivative of an expression of a function $f(x)$:

$$\frac{\partial}{\partial \boldsymbol{\phi}} \mathbb{E}_{P(x|\boldsymbol{\phi})}[f(x)]$$

with respect to the parameters $\boldsymbol{\phi}$ of the distribution $P(x|\boldsymbol{\phi})$. Show that this derivative can also be computed as:

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\phi}} \mathbb{E}_{P(x|\boldsymbol{\phi})}[f(x)] &= \mathbb{E}_{P(x|\boldsymbol{\phi})}\left[f(x)\frac{\partial}{\partial \boldsymbol{\phi}}\log(P(x|\boldsymbol{\phi}))\right] \\ &\approx \frac{1}{I}\sum_{i=1}^{I} f(x_i)\frac{\partial}{\partial \boldsymbol{\phi}}\log(P(x|\boldsymbol{\phi})).\end{aligned}$$

This method is known as the REINFORCE algorithm or score function estimator.

Problem 17.6

Why is it better to use spherical linear interpolation rather than regular linear interpolation when moving between points in the latent space? Hint: consider figure 8.13

Problem 17.7

Derive the EM algorithm for the 1D mixture of Gaussians algorithm with N components. To do this, you need to (i) find an expression for the posterior distribution $P(z|x)$ over the latent variable $z \in \{1, 2, \ldots, N\}$ for a data point x and (ii) find an expression that updates the evidence lower bound given the posterior distributions for all of the data points. You will need to use Lagrange multipliers to ensure that the weights $\lambda_1, \ldots, \lambda_N$ of the Gaussians sum to one.