cs281 exercise

Robert Gordan

December 2017

1 Importance Sampling

1.1 Problem

Derive $\mu = E(f(x))$ for an importance sampler with a mixture model, where there are K distributions q_k, each with prior probability π_k.

1.2 Solution

The equation for the importance sampler estimator is:

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\frac{f(x_i)p(x_i)}{q(x_i)}$$

To get the probability of an individual sample under our proposal distribution, we simply sum up the likelihoods under each mixure model:

$$q(x_i) = \sum_{k=1}^{K}\pi_k q_k(x_i)$$

This gives:

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\frac{f(x_i)p(x_i)}{\sum_{k=1}^{K}\pi_k q_k(x_i)}$$

2 Positivisation

2.1 Problem

We know that the optimal density for our sampling density q is:

$$q^* = \frac{|f(x)|p(x)}{\int |f(x)|p(x)dx}$$

In this case the variance of the estimator is 0 if we have $f(x) \geq 0$ or $f(x) \leq 0$. However, if $f(x)$ can take positive or negative values, the variance of the estimator under q^*, this property no longer holds. Develop a modification to standard importance sampling such that the optimal sampling density achieves this property, and show that it does so.

2.2 Solution

The key is an extension of importance sampling called multiple importance sampling. In multiple importance sampling we have multiple densities and sample from them according to a given number of times and weight each sample according to a "partition of unity" $w_j(x)$ for all densities $q_j(x)$. Like mixture sampling, this method in general helps model multimodal densities we may want in order to reach multiple peaks of f.
We need to break f into two functions:

$$f_+(x) = \max(f(x), 0)$$
$$f_-(x) = \max(-f(x), 0)$$

Then the corresponding densities are:

$$f_+(x)p(x) > 0 \Rightarrow q_+(x) > 0$$
$$f_-(x)p(x) > 0 \Rightarrow q_-(x) > 0$$

This gives us:

$$\mu = \frac{1}{n_+}\sum_{i=1}^{n_+}\frac{f_+(x_+)p(x_+)}{q_+(x_+)} - \frac{1}{n_-}\sum_{i=1}^{n_-}\frac{f_-(x_-)p(x_-)}{q_-(x_-)}$$

This is an unbiased estimator because $f_+(x) - f_-(x) = f(x)$. Taking the variance, we get:

$$Var(\mu) = E[(\frac{1}{n_+}\sum_{i=1}^{n_+}\frac{f_+(x_+)p(x_+)}{q_+(x_+)} - \frac{1}{n_-}\sum_{i=1}^{n_-}\frac{f_-(x_-)p(x_-)}{q_-(x_-)})^2]$$
$$-E[\frac{1}{n_+}\sum_{i=1}^{n_+}\frac{f_+(x_+)p(x_+)}{q_+(x_+)} - \frac{1}{n_-}\sum_{i=1}^{n_-}\frac{f_-(x_-)p(x_-)}{q_-(x_-)}]^2$$

Simplifying:

$$Var(\mu) = \frac{1}{n_+}\int\frac{(p(x)f_+(x) - \mu_+q_+(x))^2}{q_+(x)}dx + \frac{1}{n_-}\int\frac{(p(x)f_-(x) - \mu_-q_-(x))^2}{q_-(x)}$$

The two components of the sum are equivalent to those for regular importance sampling. Thus, because the two parts of our decomposed function fulfill the condition of not being both negative and positive, the expressions can both be made to be zero with the optimal choices for q_+ and q_-.