Deep Learning - Foundations and Concepts

Chapter 14. Sampling

nonlineark@github

March 25, 2025

Outline

Expectations

For some applications the goal is to evaluate expectations with respect to the distribution. Suppose we wish to find the expectation of a function $f(z)$ with respect to a probability distribution $p(z)$:

$$E(f) = \int f(z)p(z)\mathrm{d}z$$

The general idea behind sampling methods is to obtain a set of samples $z^{(l)}$ drawn independently from the distribution $p(z)$. This allows the expectation to be approximated by a finite sum:

$$\bar{f} = \frac{1}{L}\sum_{l=1}^{L} f(z^{(l)})$$

Expectations

Let's calculate the expectation and variance of $\bar{f}$:

$$E(\bar{f}) = E(\frac{1}{L}\sum_{l=1}^{L} f(z^{(l)})) = E(f)$$

$$E(\bar{f}^2) = E(\frac{1}{L^2}\sum_{l,l'} f(z^{(l)})f(z^{(l')})) = (E(f))^2 + \frac{1}{L}\mathrm{var}(f)$$

$$\mathrm{var}(\bar{f}) = E(\bar{f}^2) - (E(\bar{f}))^2 = \frac{1}{L}\mathrm{var}(f)$$

Which shows that:

- $\bar{f}$ is an unbiased estimator of $E(f)$.
- Due to the linear decrease of the variance with increasing L, in principle, high accuracy may be achievable with a relatively small number of samples $z^{(l)}$.

Standard distributions

Problem

Suppose that z is uniformly distributed over the interval $(0, 1)$. Given a probability density function p, find a function g such that the random variable $y = g(z)$ has p as its probability density function.

Standard distributions

Let U be the probability density function of the uniform distribution over the interval $(0, 1)$, we have:

$$
\begin{aligned}
p(y)\mathrm{d}y &= U(z)\mathrm{d}z \\
f(y_0) &= \int_{-\infty}^{y_0} p(y)\mathrm{d}y = \int_{-\infty}^{z_0} U(z)\mathrm{d}z = z_0 \\
y_0 &= f^{-1}(z_0)
\end{aligned}
$$

So we have to transform the uniformly distributed random numbers using a function that is the inverse of the cumulative distribution function of the desired probability density function.

Standard distributions

Some examples:

- Exponential distribution $p(y) = \lambda \exp(-\lambda y)$:
  - $z = f(y) = \int_0^y p(t)\mathrm{d}t = 1 - \exp(-\lambda y)$.
  - $y = -\frac{1}{\lambda} \log(1 - z)$.
- Cauchy distribution $p(y) = \frac{1}{\pi} \frac{1}{1+y^2}$:
  - $z = f(y) = \int_{-\infty}^y p(t)\mathrm{d}t = \frac{1}{\pi} \arctan y + \frac{1}{2}$.
  - $y = \tan(\pi(z - \frac{1}{2}))$.

Standard distributions

The generalization to multiple variables involves the Jacobian of the change of variables, so that:

$$p_Y(y_1, \ldots, y_M) = p_Z(z_1, \ldots, z_M)\left|\frac{\partial(z_1, \ldots, z_M)}{\partial(y_1, \ldots, y_M)}\right|$$

Standard distributions

The Box-Muller method for generating samples from a Gaussian distribution. First, suppose we generate pairs of uniformly distributed random numbers $z_1, z_2 \in (-1, 1)$. Next, we discard each pair unless it satisfies $z_1^2 + z_2^2 \leq 1$. This leads to a uniform distribution of points inside the unit circle with $p_Z(z_1, z_2) = \frac{1}{\pi}$. Then, for each pair z_1, z_2 we evaluate the quantities:

$$y = z\frac{\sqrt{-4\log||z||}}{||z||}$$

The joint distribution of y_1 and y_2 is given by:

$$p_Y(y_1, y_2) = p_Z(z_1, z_2)|\frac{\partial(z_1, z_2)}{\partial(y_1, y_2)}| = (\frac{1}{\sqrt{2\pi}}\exp(-\frac{y_1^2}{2}))(\frac{1}{\sqrt{2\pi}}\exp(-\frac{y_2^2}{2}))$$

So y_1 and y_2 are independent and each has a Gaussian distribution with zero mean and unit variance.