COMP9444 Neural Networks and Deep Learning

Quiz 2 (Probability and Backprop Variations)

This is an optional quiz to test your understanding of the material from Week 2.

---

1. Write the formula for a Gaussian distribution with mean μ and standard deviation σ.

   $P(x) = \exp(-(x-\mu)^2/2\sigma^2) / (\mathrm{sqrt}(2\pi)\sigma)$

2. Write the formula for Bayes' Rule, in terms of a cause A and an effect B.

   $P(A|B) = P(B|A)P(A) / P(B)$

3. Write the formula for the Entropy H(p) of a continuous probability distribution p()

   $H(p) = \int_\theta p(\theta)\ (-\log p(\theta))\ d\theta$

4. Write the formula for the Kullback-Leibler Divergence $D_{KL}(p \,||\, q)$ between two continuous probability distributions p() and q().

   $D_{KL}(p \,||\, q) = \int_\theta p(\theta)\ (\log p(\theta) - \log q(\theta))\ d\theta$

5. Write the formulas for these Loss functions: Squared Error, Cross Entropy, Weight Decay. (remember to define any variables you use)

   Assume z_i is the actual output, t_i is the target output and w_j are the weights.
   Squared Error: $E = ½ \Sigma_i (z_i - t_i)^2$
   Cross Entropy: $E = \Sigma_i (-t_i \log(z_i) - (1- t_i)\log(1- z_i))$
   Weight Decay: $E = ½ \Sigma_j w_j^2$

6. In the context of Supervised Learning, explain the difference between Maximum Likelihood estimation and Bayesian Inference.

In Maximum Likelihood estimation, the hypothesis $h \in H$ is chosen which maximizes the conditional probability $P(D \mid h)$ of the observed data D, conditioned on h. In Bayesian Inference, the hypothesis $h \in H$ is chosen which maximizes $P(D \mid h)P(h)$, where $P(h)$ is the prior probability of h.

7. Briefly explain the concept of Momentum, as an enhancement for Gradient Descent.

A running average of the differentials for each weight is maintained and used to update the weights as follows:

$$\delta w = \alpha \delta w - \eta \, dE/dw$$
$$w = w + \delta w$$

The constant α with $0 \leq \alpha < 1$ is called the momentum.
