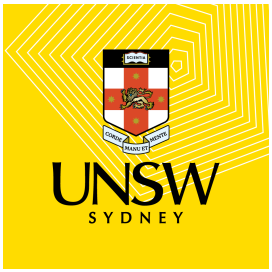
COMP9444
Neural Networks and Deep Learning
Term 2, 2024



Week 3 Tutorial: Probability, Generalisation and Overfitting

1. **Bayes' Rule**

   One bag contains 2 red balls and 3 white balls. Another bag contains 3 red balls and 2 green balls. One of these bags is chosen at random, and two balls are drawn randomly from that bag, without replacement. Both of the balls turn out to be red. What is the probability that the first bag is the one that was chosen?

2. **Entropy and KL-Divergence for Discrete Distributions**

   Consider these two probability distributions on the same space $\Omega = \{A, B, C, D\}$

   $p = \langle \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8} \rangle$
   $q = \langle \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2} \rangle$

   (a) Construct a Huffman tree for each distribution p and q
   (b) Compute the entropy $H(p)$
   (c) Compute the KL-Divergence in each direction $D_{KL}(q||p)$ and $D_{KL}(p||q)$. Which one is larger? Why?

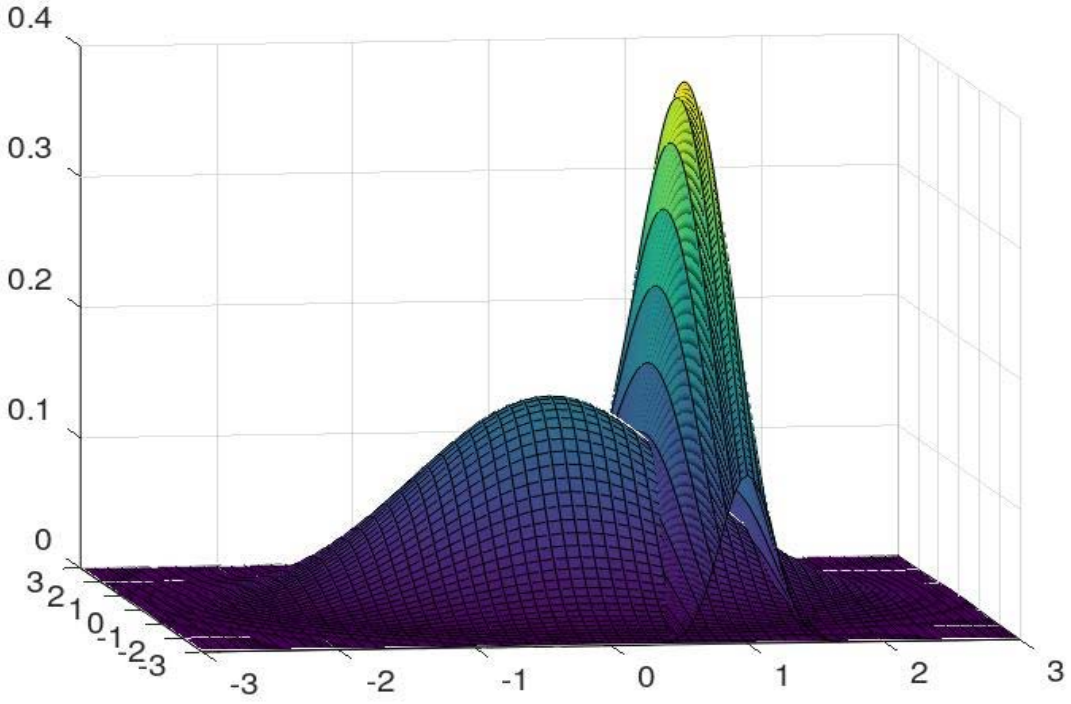
3. **Entropy, KL-Divergence and W_2 Distance for Bivariate Gaussians**

   Consider two bivariate Gaussian distributions p and q (see figure).

   q has mean $\mu_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and variance $\sum_1 = \begin{bmatrix} 0.04 & 0 \\ 0 & 4 \end{bmatrix}$

   p has mean $\mu_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and variance $\sum_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

   (a) Compute the Entropy $H(p)$ and $H(q)$. Which one is larger? Why?
   (b) Compute the KL-Divergence in each direction $D_{KL}(q||p)$ and $D_{KL}(p||q)$. Which one is larger? Why?
   (c) Compute the Wasserstein Distance $W_2(q, p)$



4. **Any Other Questions**

   Any further questions or discussion about PyTorch, other parts of the course, or broader implications of deep learning.