CSC2516: Homework #3

Due on March 07

Roger Grosse, Jimmy Ba

Thomas Hollis

Problem 1

Dropout has an interesting interpretation in the case of linear regression. Find expressions for $\mathbb{E}[y]$ and $\mathrm{Var}[y]$ for a given $\mathbf{x}$ and $\mathbf{w}$.

Solution

From the stochastic prediction equation we can write that the expectation of $\mathbf{y}$ for a given $\mathbf{x}$ and $\mathbf{w}$ is:

$$\mathbb{E}[y] = \mathbb{E}\big[\sum_j m_j w_j x_j\big] = \sum_j x_j w_j \mathbb{E}\big[m_j\big] \tag{1}$$

Since the Bernoulli random variable has an expectation of 0.5 we can therefore state:

$$\mathbb{E}[y] = 0.5 \sum_j x_j w_j \tag{2}$$

From the stochastic prediction equation we can write that the variance of $\mathbf{y}$ for a given $\mathbf{x}$ and $\mathbf{w}$ is:

$$\mathrm{Var}[y] = \mathrm{Var}\big[\sum_j m_j w_j x_j\big] = \sum_j x_j^2 w_j^2 \mathrm{Var}\big[m_j\big] \tag{3}$$

Since the Bernoulli random variable has a variance of 0.25 we can therefore state:

$$\mathrm{Var}[y] = 0.25 \sum_j x_j^2 w_j^2 \tag{4}$$

Problem 2

Determine $\tilde{y}_j$ as a function of w_j such that $\mathbb{E}[y] = \tilde{y} = \sum_j \tilde{w}_j x_j$.

Solution

By equating the above equality with (2), we can state:

$$\mathbb{E}[y] = \tilde{y} = \sum_j \tilde{w}_j x_j = 0.5 \sum_j x_j w_j \tag{5}$$

Hence by rearranging (5) and because the product of a sum is the sum of products:

$$\sum_j \tilde{w}_j x_j = \sum_j (0.5 \times w_j) x_j \tag{6}$$

Thus we can conclude from (6) that:

$$\tilde{w}_j = 0.5 \times w_j \tag{7}$$

Problem 3

Using the model from the previous section, show that the cost J can be written as $J = \frac{1}{2N}\sum_{i=1}^{N}(\tilde{y}^{(i)} - t^(i))^2 + R(\tilde{w}_1, \ldots, \tilde{w}_D)$.

Solution

From the cost function stated in the handout, we have:

$$J = \frac{1}{2N}\sum_{i=1}^{N}\mathbb{E}[(y^{(i)} - t^{(i)})^2] \tag{8}$$

By expanding (8) we have:

$$J = \frac{1}{2N}\sum_{i=1}^{N}\mathbb{E}[(y^{(i)})^2] - 2\mathbb{E}[y^{(i)}t^{(i)}] + \mathbb{E}[(t^{(i)})^2] \tag{9}$$

Simplifying (9) in turn yields:

$$J = \frac{1}{2N}\sum_{i=1}^{N}\mathbb{E}[(y^{(i)})^2] - 2t^{(i)}\mathbb{E}[y^{(i)}] + (t^{(i)})^2 \tag{10}$$

Using properties of expectation and variance we can further reduce (10) to:

$$J = \frac{1}{2N}\sum_{i=1}^{N}\mathbb{E}[(y^{(i)}]^2 + \mathrm{Var}[y^{(i)}] - 2t^{(i)}\mathbb{E}[y^{(i)}] + (t^{(i)})^2 \tag{11}$$

Thus, through the equality in Problem 2, we have:

$$J = \frac{1}{2N}\sum_{i=1}^{N}(\tilde{y}^{(i)})^2 + \mathrm{Var}[y^{(i)}] - 2t^{(i)}\tilde{y}^{(i)} + (t^{(i)})^2 \tag{12}$$

Factorising (12) to attempt to get to the desired form yields:

$$J = \frac{1}{2N}\sum_{i=1}^{N}(\tilde{y}^{(i)} - t^{(i)})^2 + \frac{1}{2N}\sum_{i=1}^{N}\mathrm{Var}[y^{(i)}] \tag{13}$$

In turn, substituting (4) into (13):

$$J = \frac{1}{2N}\sum_{i=1}^{N}(\tilde{y}^{(i)} - t^{(i)})^2 + \frac{1}{2N}\sum_{i=1}^{N}0.25\sum_{j}(x_j^{(i)})^2 w_j^2 \tag{14}$$

Finally, substituting (7) into (14) yields:

$$J = \frac{1}{2N}\sum_{i=1}^{N}(\tilde{y}^{(i)} - t^{(i)})^2 + \frac{1}{2N}\sum_{i=1}^{N}\sum_{j}(x_j^{(i)})^2 \tilde{w}_j^2 \tag{15}$$

Thus, since the second term of J in (15) is a positive function of $\tilde{w}_D$ that depends on the data $\mathbf{x}$, we have shown as required that it acts as a data-dependent regulariser as follows:

$$J = \frac{1}{2N}\sum_{i=1}^{N}(\tilde{y}^{(i)} - t^{(i)})^2 + R(\tilde{w}_1, \ldots, \tilde{w}_D) \qquad \textit{where} \qquad R(\tilde{w}_1, \ldots, \tilde{w}_D) = \frac{1}{2N}\sum_{i=1}^{N}\sum_{j}(x_j^{(i)})^2 \tilde{w}_j^2 \tag{16}$$

Problem 4

Design the weights and biases for an RNN which has two input units, three hidden units, and one output unit, which implements binary addition. All of the units use the hard threshold activation function.

Solution

The RNN that needs to be designed is governed by the following equations:

$$y^{(t)} = \mathbf{v}\mathbf{h}^{(t)} + b_y \tag{17}$$

$$\mathbf{h}^{(t)} = \mathbb{1}\Big[\mathbf{U}\mathbf{x}^{(t)} + \mathbf{b_h} + \mathbf{W}\mathbf{h}^{(t-1)}\Big] \qquad \textit{where} \qquad \mathbb{1} \text{ is the unit step function} \tag{18}$$

We also know that our example input is:

$$\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \tag{19}$$

An RNN solution which achieves binary addition via the format of (17) and (18) is as follows:

$$\mathbf{U} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\mathbf{b_h} = \begin{bmatrix} -0.5 \\ -1.5 \\ -2.5 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} 1 & -1 & 1 \end{bmatrix}$$

$$b_y = -0.5$$