ID 2088192

Question 32212

After inserting your student ID and the question number in the title, header and footer, write your answers between here and the statement of good academic conduct. Your ID and the question number will automatically appear on any subsequent pages.

Q1:
(a) No.

(b) We assume N is normally distributed with mean 0 and variance σ^2.

(c) Given cost function: $J(\mathbf{w}) := \frac{1}{n}\sum_{j=1}^{n}(\mathbf{w}\cdot\mathbf{x}^{(j)} - y^{(j)})^2 + \alpha\|w\|_2$, where $\|\mathbf{w}\|_2 = \sqrt{\mathbf{w}\cdot\mathbf{w}}$ is the Euclidean norm, we have:

$$\begin{aligned}
\nabla J(\mathbf{w}) &= \frac{2}{n}\sum_{j=1}^{n}(x^{(j)}\mathbf{w} - y^{(j)})\cdot x^{(j)} + \alpha\cdot 2\mathbf{w}\cdot\frac{1}{2\sqrt{\mathbf{w}\cdot\mathbf{w}}} \\
&= \frac{2}{n}\sum_{j=1}^{n}[(x^{(j)})^2\cdot\mathbf{w} - x^{(j)}y^{(j)}] + \alpha\frac{\mathbf{w}}{\|\mathbf{w}\|_2} \\
&= [\frac{2}{n}\sum_{j=1}^{n}(x^{(j)})^2 + \frac{\alpha}{\|\mathbf{w}\|_2}]\mathbf{w} - \frac{2}{n}\sum_{j=1}^{n}x^{(j)}y^{(j)}
\end{aligned}$$

This equation is not linear, thus cannot be solved directly.

One can set the model parameters using gradient descent:

Initiate with parameter **w**, learning rate e>0.
Repeat:

$$w := w - e\cdot\nabla J(w)$$

(d) We update $\nabla J(\mathbf{w})$ using the given conditions that $y^{(j)} - \mathbf{w} \cdot \mathbf{x}^{(j)} = 0$ except $j = 1$:

$$\nabla J(\mathbf{w}) = \frac{2}{n} \sum_{j=1}^{n} (x^{(j)}\mathbf{w} - y^{(j)}) \cdot x^{(j)} + \alpha \frac{\mathbf{w}}{\|\mathbf{w}\|_2}$$
$$= \frac{2}{n}(x^{(1)}\mathbf{w} - y^{(1)}) \cdot x^{(1)} + \alpha \frac{\mathbf{w}}{\|\mathbf{w}\|_2}$$
$$= \frac{2}{n}(-\delta \cdot x^{(1)} + \frac{\alpha n}{2\|\mathbf{w}\|_2})$$

According to the updating rule of gradient descent:

$$w := w - e \cdot \nabla J(w)$$

To increase parameter w_1, given learning rate > 0, we need $\nabla J(w_1) < 0$. Thus:

$$\nabla J(w_1) = \frac{2}{n}(-\delta \cdot x^{(1)} + \frac{\alpha n}{2\|\mathbf{w}\|} w_1) < 0$$
$$\frac{\alpha n}{2\|\mathbf{w}\|} w_1 < \delta \cdot x^{(1)}$$

Given $w_1 > 0$:

$$\alpha < \frac{2\delta \cdot x^{(1)} \cdot \|\mathbf{w}\|_2}{n w_1}$$

Since $\delta > 0,\ x^{(1)} > 0,\ n \in N^+$, we verify that $\alpha > 0$. Thus:

For $\alpha \in (0, \frac{2\delta \cdot x^{(1)} \cdot \|\mathbf{w}\|_2}{n w_1})$, the model parameter w_1 will increase.

Q2:

(a) The underlying idea of using "dropout" is to approximate the ensemble methods on a large number of neural networks to reduce generalisation error.

(b) During training of the neural network. It implicitly averages outputs from all subnetworks to learn the weights for the full network. When testing or using the network, the full neural network should be used, otherwise will return stochastic results from some subnetworks.

(c) Given dropout rate p unchanged, define a noise term $\epsilon \sim \mathcal{N}(1, \sigma^2)$ on the probability of the Bernoulli distribution $d_j^l \sim Ber(1-p)$, redefine the dropout activation unit as:

$$\tilde{a}_j^l = \frac{1}{1 - p \cdot \epsilon} \cdot Ber(1 - p \cdot \epsilon) \cdot \phi(z_j^l)$$

One can show that the expectation of the activation unit stays:

$$\mathbb{E}[\tilde{a}_j^l] = p\epsilon \cdot \frac{1}{1 - p\epsilon} \cdot 0 \cdot \phi(z_j^l) + (1 - p\epsilon) \cdot \frac{1}{1 - p\epsilon} \cdot 1 \cdot \phi(z_j^l)$$
$$= 0 + \phi(z_j^l) = a_j^l$$

Do not write below this line

Statement of good academic conduct

By submitting this assignment, I understand that I am agreeing to the following statement of good academic conduct:

- I confirm that this assignment is **my own work** and I have not worked with others in preparing this assignment.
- I confirm this assignment was written by me and is in my own words, except for any materials from published or other sources which are clearly indicated and acknowledged as such by appropriate referencing.
- I confirm that this work is not copied from any other person's work (published or unpublished), web site, book or other source, and has not previously been submitted for assessment either at the University of Birmingham or elsewhere.
- I confirm that I have not asked, or paid, others to prepare any part of this work for me.
- I confirm that I have read and understood the University regulations on plagiarism (https://intranet.birmingham.ac.uk/as/registry/policy/conduct/plagiarism/index.aspx).