CS5370: Assignment 2

Nakka Chakradhar - EE16BTECH1102

September 22, 2019

1. Settting some parameters :

   Let input be a D dimension vector.

   Let output be a K dimensional vector

   Let the loss function

   s_{ji} denote the skip layer weight between i^{th} input and the j^{th} output.

   a_i^l denotes the output of the i^{th} neuron in l^{th} layer. This layer has M nodes for reference (Given 2 in the problem).

   σ is the activation function in the last layer of neurons.

   gt is the target output.

   $output_k^2 = \sigma(\Sigma_{i=0}^{M} w_{ki}^l a_i^l + \Sigma_{i=0}^{D} s_{ki} x_i)$

   Let $z_k^2 = \Sigma_{i=0}^{M} w_{ki}^l a_i^l + \Sigma_{i=0}^{D} s_{ki} x_i$

   Then, $a_k^2 = \sigma(z_k^2)$

   E $= \frac{1}{2}\Sigma_{i=1}^{K}(a_i^2 - gt_i)^2$, for one training example.

   $\frac{\partial E}{\partial s_{ji}} = (a_j^2 - gt_j) \ \times \ \sigma'(z_j^2) \ \times \ x_i$

2. Answer to question 2

   $E(w) \ \approx \ E(\tilde{w}) + (w - \tilde{w})^T \nabla E|_{w=\tilde{w}} + \frac{1}{2}(w - \tilde{w})^T H (w - \tilde{w})$

   From the above expression, it is clear that the first term is a constant and the second term has W partial derivatives.

   Since Hessian is symmetric, it has $\frac{\partial E}{\partial w_i \partial w_j} = \frac{\partial E}{\partial w_j \partial w_i}, \ \forall i, j \in \{1, 2, ..., W\}$.

   Number of independent elements from Hessian $= \frac{W(W-1)}{2} + W$

   Number of independent elements from $(w - \tilde{w}) = W$

   Total $= \frac{W(W-1)}{2} + 2W = \frac{W(W+3)}{2}$

3. Answer to question 3

   Given : $x = \Sigma_i \alpha_i x_i$

   $\alpha_i >= 0$

   $\Sigma \alpha_i = 1$

   $w^T x_i + w_0 > 0$

$z = \Sigma_i \beta_i z_i$

$\beta_i >= 0$

$\Sigma \beta_i = 1$

$w^T z_i + w_0 < 0$

1. The convex hulls intersect at a point say A. A belongs to both the convex hulls. $\exists$ unique α and β such that A = $\Sigma_i \alpha_i x_i = \Sigma_i \beta_i z_i$.

Let's assume that the points are linearly separable and prove the statement via contradiction

Let's substitute this point in the equation of the decision boundary.

Since sum of αs, βs is 1, we have

$f(A) = \Sigma_i (w^T \alpha_i x_i) + w_0 = \Sigma_i \alpha_i (w^T x_i + w_0) = \Sigma_i \beta_i (w^T z_i + w_0)$

Since the points lie on different sides of the line, we have f(A) $>$ 0 and f(A) $<$ 0 simultaneously. This is a contradiction.

2. Given the convex hulls are linearly separable. Assume they are intersecting at a point A. A belongs to both the convex hulls. $\exists$ unique αs and βs such that A = $\Sigma_i \alpha_i x_i = \Sigma_i \beta_i z_i$.

Let's substitute this point in the equation of the decision boundary.

Since sum of alphas is 1, we have

$f(A) = \Sigma_i (w^T \alpha_i x_i) + w_0 = \Sigma_i \alpha_i (w^T x_i + w_0) = \Sigma_i \beta_i (w^T z_i + w_0)$

Since the points lie on different sides of the line, we have f(A) $>$ 0 and f(A) $<$ 0 simultaneously. This is a contradiction. And hence our assumption that there exists a point A which is in common for both the convex hulls is false.

(The proofs are pretty similar because they are based on one assumption - there is a point in common and that point won't follow standard comparison methods leading to a contradiction).