RBFN

Neal Moorthy

4/2/18

1. If we take the r 1-4 as our 4 basis vectors, then the weight matrix is assigning the classifiers associated with each basis vector and a bias of zero because the data is already centered around the origin. In the weight vector r1 and r2 are 1 because their two inputs match, while r3 and r4 are -1 because their inputs don't match. If we plug these basis vectors into the ϕ(x) equation from the lecture notes (exp(x-ri)) then we get W.

2. In a multi-class classification setting one could perform a nearest neighbor classification by RBFN by taking each input vector of size K (dimension of input space), appending the correct classification (K+1), transposing to make each column a basis vector and we have L basis vectors (number of input vectors). This [K+1][L] matrix could then be used to perform nearest neighbor classification by computing the distance of the new input vector to each basis vector.

3. $D(y^*, M, \phi(x)) = -(y^* \log(\sigma(w^T\phi(x)) + (1-y^*)\log(1-\sigma(w^T\phi(x)))$

First, we derive the derivative with respect to $\phi_k(x)$

$\nabla_{\phi_k(x)} D(y^*, M, \phi(x))$

$$= (-y^*)(\frac{1}{\sigma(w^T\phi(\tilde{x}))})(\sigma(w^T\phi(\tilde{x}))\left(1-\sigma(w^T\phi(\tilde{x}))\right)w_k - (1 - y^*)(\frac{1}{\sigma(w^T\phi(\tilde{x}))})(-1)(\sigma(w^T\phi(\tilde{x}))\left(1-\sigma(w^T\phi(\tilde{x}))\right))w_k$$

$$\nabla_{\phi_k(x)} D(y^*, M, \phi(x)) = -y^*\left(1-\sigma(w^T\phi(\tilde{x}))\right)w_k + (1-y^*)\left(1-\sigma(w^T\phi(\tilde{x}))\right)w_k$$

$$\nabla_{\phi_k(x)} D(y^*, M, \phi(x)) = [-y^* + y^*\sigma(w^T\phi(\tilde{x})) + \sigma(w^T\phi(\tilde{x})) - y^*\sigma(w^T\phi(\tilde{x}))]w_k$$

$$\nabla_{\phi_k(x)} D(y^*, M, \phi(x)) = -(y^* - \sigma(w^T\phi(\tilde{x})))w_k$$

Next, we take a look a $\phi_k(x) = \exp(-(x-r^k)^2)$

Solving for $\frac{\partial\phi_k(x)}{\partial r^k}$ allows us to use the chain rule to find $\nabla_{\phi_k(x)} D(y^*, M, \phi(x))$.

$$\frac{\partial\phi_k(x)}{\partial r^k} = \exp(-(x-r^k)^2) * \left(-2(x-r^k)\right) * (-1)$$

$$\frac{\partial\phi_k(x)}{\partial r^k} = 2 * \exp(-(x-r^k)^2) * (x-r^k)$$

Therefore,

$$\nabla_{r^k} D(y^*, M, \phi(x)) = -(y^* - \sigma(w^T\phi(\tilde{x})))w_k * 2 * \exp(-(x-r^k)^2) * (x-r^k)$$

And if we re-write the equation and substitute back in $\phi_k(x)$ matches the equation in our notes:

$$\nabla_{r^k} D(y^*, M, \phi(x)) = -2(y^* - \sigma(w^T\phi(\tilde{x})))w_k * \phi_k(x) * (x-r^k)$$