

Lecture 8

Name: Harsh Sanjay Roniyar

Roll Number: 22B3942

Continuing from where we left in the last lecture...

The nature of the bias-variance relations can be better visualized using the contour plots of y , t , and x .

Now, the calculation of the expected loss, as left in [Lecture 7](#) becomes -

$$\mathbb{E}[L] = \int \{y(\mathbf{x}) - \mathbb{E}[t|x]\}^2 p(x) dx + \int \{\mathbb{E}[t|x] - t\}^2 p(x) dx$$

The second term in this loss, can be related to the irreducible noise of our data. We can attempt to reduce the first part of the expected loss from our end.

$$h(x) = \mathbb{E}[t|x] = \int t p(t|x) dt$$

Now, $\{y(x) - h(x)\}$ is meaningless since our $y(x)$ depends on the data samples. Hence, we will represent it using $\{y(x; D) - h(x)\}^2$ where D is the sample of the distribution.

$$= \{y(x; D) - \mathbb{E}_D[y(x; D)] + \mathbb{E}_D[y(x; D)] - h(x)\}^2,$$

where $\mathbb{E}_D[y(x; D)]$ is the expectation over the model. $h(x)$ can be seen as the best theoretical model, and $y(x; D)$ is the particular model under consideration for the given training data.

On further simplification, the final expression reduces to -

$$\mathbb{E}_D [\{y(x; D) - h(x)\}^2] = \underbrace{\{\mathbb{E}_D[y(x; D)] - h(x)\}^2}_{(\text{bias})^2} + \underbrace{\mathbb{E}_D[\{y(x; D) - \mathbb{E}_D[y(x; D)]\}^2]}_{\text{variance}}$$

Hence, the expected loss can be written as $(\text{bias})^2 + \text{variance} + \text{noise}$. The noise relation will be

$$\text{noise} = \int \{h(x) - t\}^2 p(x, t) dx dt$$

Model Regularization

L2 Regularization

$$y(x_i) = \sum_{j=1}^D w_j x_{ij} + w_0$$

Then, the L_p norm of \mathbf{w} is given by

$$\left(\sum_{j=1}^D |w_j|^p \right)^{\frac{1}{p}}$$

The norms of $L_\infty = \max |w_j|$ and $L_0 = \text{count of non-zero components}$.

The objective function is -

$$\min_w \frac{1}{2} \sum_{i=1}^N (y(x_i) - t_i)^2 + \frac{\lambda}{2} \sum_{j=1}^D |w_j|^q$$

with,

$$\sum_{j=1}^D |w_j|^q \leq \eta$$

The objective can be further written as -

$$\begin{aligned} \min_{\mathbf{w}} \frac{1}{2} \sum_{i=1}^N (t_i - \mathbf{w}^T \phi(\mathbf{x}_n))^2 + \frac{\lambda}{2} \mathbf{w}^T \mathbf{w} \\ \implies E_D(\mathbf{w}) + \lambda E_W(\mathbf{w}). \end{aligned}$$

The optimized solution for \mathbf{w} is -

$$\boxed{\mathbf{w} = (\lambda \mathbf{I} + \Phi^T \Phi)^{-1} \Phi^T \mathbf{t}}.$$

The dimension of Φ is $N \times D$. Hence, we are taking the pseudo-inverse.