Lecture 7

Name: Harsh Sanjay Roniyar

Roll Number: 22B3942

Linear Regression

Data: $[x_1 \dots x_D]$, t
Predictions: $y = f_w(\bar{x})$
where, $t_i, y_i \in \mathbb{R}, \forall i$

$$y_i = \left(\sum_{j=1}^{D} w_j x_{ij}\right) + w_0 = W^T x$$

This acts as a linear approximation of our data.

Since $x_i \in \mathbb{R}$. Then I can create $D+1$ features in $\phi_i = \begin{bmatrix} 1 \\ x_i \\ x_i^2 \\ \vdots \\ x_i^D \end{bmatrix}$. Then this is a polynomial representation in the x-space, but in linear form in the space ϕ.

For (all) the data-points, fitting a Gaussian curve at those points and then adding the distributions also gives us a fit for the data.

$$\phi = \left[e^{-\frac{(x-x_j)^2}{2\sigma^2}}\right]$$

where ϕ is a $W \times N$ matrix which can be reduced to $W \times D$.

Radial Basis Function. Similar to Kernel Density Estimation.

Now, back to linear regression.
The predictions t, would have some stochastic noise.

$$t = y(\mathbf{x}, \mathbf{w}) + \epsilon$$

Assuming the noise term to be Gaussian and $\beta = \frac{1}{\sigma^2}$

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{i=1}^{N} N(t_i|\mathbf{w}^T \phi(x_i), \beta^{-1})$$

Now, since this has product terms, taking $\log$:

$$\ln p(\mathbf{t}|\mathbf{w}, \beta) = \sum_{i=1}^{N} \ln \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{\{t_i - \mathbf{w}^T \phi(x_i)\}^2}{2\sigma^2}$$

$$E_D(w) = \frac{1}{2} \sum_{i=1}^{N} \{t_i - \mathbf{w}^T \phi(x_i)\}^2$$

Now, going on to maximizing the likelihood - for this we need to set the gradients w.r.t x to zero. So, we are maximizing,

$$\nabla \ln p(\mathbf{t}|\mathbf{w}, \beta) = \sum_{i=1}^{N} \{t_n - \mathbf{w}^T \phi(x_n)\} \phi(x_n)^T$$

which finally gives,

$$\mathbf{w_{ML}} = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \mathbf{t}$$

where,

$$\mathbf{\Phi} = \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_N)^T \end{bmatrix}$$

Conclusions:

1. Assume linear model
2. Assume error is Gaussian distributed
   1. M.S.E is the metric to maximize data likelihood
   2. The analytical solution for $w = \text{PseudoInv}(\mathbf{w}) \times \mathbf{t}$

Bias-Variance Decomposition

`of linear regression and gaussian noise`

$$L_i = \text{Loss} = (y(x_i) - t_i)^2$$

Then,

$$\mathbb{E}(L) = \int \int L \cdot p(x, t) dx dt = \int \int (y(x) - t)^2 p(x, t) dx dt$$

$$\frac{\delta \mathbb{E}(L)}{\delta y(x)} = 2\int \{y(x) - t\} p(x,t) dt = 0$$

This gives,

$$y(x) = \frac{\int t p(x,t) \mathrm{d}t}{p(x)} = \mathbb{E}_t(t|x)$$

Now, if we try to do manipulation with

$$\{y(x) - t\}^2 = \{y(x) - \mathbb{E}(t|x) + \mathbb{E}(t|x) - t\}^2$$

$$= \{y(x) - \mathbb{E}(t|x)\}^2 + \{\mathbb{E}(t|x) - t\}^2 + 2\{y(x) - \mathbb{E}(t|x)\}\{\mathbb{E}(t|x) - t\}$$