Model Selection

Dr. Kalidas Y., IIT Tirupati

In this lecture, you will understand that multiple types of models is possible to have

Can you have 1000 models?

$$X \times w = y$$

$$? \qquad\qquad ? \qquad\qquad ?$$

Can you have 1000 models?

$$X \times w = y$$

?   ?   ?

- Determining a w such that $L(w)$ is minimized
  Symbolically denoted as,

$$w^* = \underset{w}{\operatorname{argmin}}\, L(w)$$

- NOTE: $\min_{w} L(w)$ is different from $\underset{w}{\operatorname{argmin}}\, L(w)$
- $\min_{w} L(w)$ means **minimum value** of Loss function across various values of w
- $\underset{w}{\operatorname{argmin}}\, L(w)$means **minimizing vector** w for a given Loss function

$$L(w) = \sum_{i=1}^{i=N} (y_i - w \cdot x_i)^2$$

this is an implicit form.. “data set” is assumed to have been given

explicitly compute loss function value on a given “data set”

data set

$$L(w, D) = \sum_{(x,y) \in D} (y - x \cdot w)^2$$

point in a data set

each point in a data set is a tuple

Can you have 1000 models?

$$w_1^* = \underset{w}{\operatorname{argmin}}\, L(w, D_1)$$

$$w_2^* = \underset{w}{\operatorname{argmin}}\, L(w, D_2)$$

…

…

$$w_{1000}^* = \underset{w}{\operatorname{argmin}}\, L(w, D_{1000})$$

"Build a model" FOR EACH "data set"

Can you have 1000 models?

$$X \times w = y$$

$$? \qquad ? \qquad ?$$

Fitting a Line Passing Through Origin

- $y = m\,x$
- $L(m) = \sum_{i=1}^{i=N}(y_i - m\,x_i)^2$
- $X = \begin{bmatrix} x_1 \\ \cdots \\ x_N \end{bmatrix}_{N\times 1}$, $Y = \begin{bmatrix} y_1 \\ \cdots \\ y_N \end{bmatrix}_{N\times 1}$, $W = [m\,]_{1\times 1}$
- $L([m]) = (XW - Y)^T(XW - Y)$
- $\nabla L = \left[\frac{\partial L}{\partial m}\right]$ //It's a function
- $W_{(new)} = W_{(old)} - \nabla L|_{W=W_{(old)}}$

Squared error type loss function

Fitting a Line – slope and intercept

- $y = m\,x + c$
- $L(m) = \sum_{i=1}^{i=N}(y_i - (m\,x_i + c))^2$
- $X = \begin{bmatrix} x_1 & 1 \\ \dots & \\ x_N & 1 \end{bmatrix}_{N\times 2}$, $Y = \begin{bmatrix} y_1 \\ \dots \\ y_N \end{bmatrix}_{N\times 1}$,

  $W = \begin{bmatrix} m \\ c \end{bmatrix}_{2\times 1}$
- $L\left(\begin{bmatrix} m \\ c \end{bmatrix}\right) = (XW - Y)^T(XW - Y)$

- $\nabla L = \begin{bmatrix} \frac{\partial L}{\partial m} \\ \frac{\partial L}{\partial c} \end{bmatrix}$ //It's a function
- $W_{(new)} = W_{(old)} - \nabla L|_{W=W_{(old)}}$

Squared error type loss function

Fitting a Parabola?

- $y = a\,x^2 + b\,x + c$
- $L(a, b, c) = \sum_{i=1}^{i=N} \big(y_i - (a\,x_i^2 +$

Squared error type loss function

Fitting a Cubic curve?

- $y = a\,x^3 + b\,x^2 + c\,x + d$
- $L(m) = \sum_{i=1}^{i=N}\big(y_i - (a\,x_i^3 + b\,x_i^2 +$

Squared error type loss function

Fitting a Degree-K polynomial?

- $y = a_k\, x^k + \cdots + a_0$
- $L(m) = \sum_{i=1}^{i=N} \left(y_i - \sum_{j=0}^{k} a_j\, x^j \right)^2$
- $X = \begin{bmatrix} x_1^k \ldots x_1^2 & x_1^1 & 1 \\ & \ldots & \\ x_N^k \ldots x_N^2 & x_N^1 & 1 \end{bmatrix}_{N \times (k+1)}$,
- $Y = \begin{bmatrix} y_1 \\ \ldots \\ y_N \end{bmatrix}_{N \times 1}$,
- $W = \begin{bmatrix} a_0 \\ \ldots \\ a_k \end{bmatrix}_{(k+1) \times 1}$
- $L(W) = (XW - Y)^T (XW - Y)$
- $\nabla L = \begin{bmatrix} \frac{\partial L}{\partial a_0} \\ \ldots \\ \frac{\partial L}{\partial a_k} \end{bmatrix}$ //It's a function
- $W_{(new)} = W_{(old)} - \nabla L|_{W = W_{(old)}}$

Squared error type loss function

Can you have 1000 models?

- Build Degree 0 Polynomial
- Build Degree 1 Polynomial
- Build Degree 2 Polynomial
- …
- …
- …
- Build Degree k polynomial

"Build a model" FOR EACH "degree"
ON A GIVEN single "data set"

Multiple Models

1. Either on different data sets
2. Or on different parameter settings of a type of model

   For example,

   - Type of model = Polynomial fitting type
   - Parameter setting = Degree of the polynomial
3. Or different types of models