Lecture 1 Jan 14 th

What is Machine Learning?

It's an algorithm that has a specific goal; solving given task

Not all tasks can be solved with an explicit algorithm.

Thus we use ML to solve these tasks by traing models using data.

3 Main Ways of ML

1. Supervised Learning
   - humans have annotated the data
     i.e. the correct data has been provided
   - Ex, classification

2. Unsupervised Learning
   - no annotations, ex. grouping

3. Reinforced Learning
   - provide incentive for completing task
   - provide punishment for screwing up task

Appraoching ML Problems

1. What's the best technique to solve problem
2. Where can I get my data
3. Select an appropiate model
4. Training the model
5. Tuning the model

- see how it works on new data

Problems in Data

- where to get the data
- no variation in data
- overfitting, model not generalized
  - to fix this problem, we split avaliable data into 3 sets:
  - training, tuning process
  - validation, evaluating model
  - testing, final test using unbias data.
- not enough data

Regression

Regression is fitting model to given data, avoiding noise
It's useful for interpolation/extrapolation.

1D Input 1D Output

We solve:

$$y = wx + b$$

for input x, output y and tuning params w, b.

We evalute this function using a Loss Function / Objective Function

For above $y = wx + b$ one L.F. we can use is:

Lease Square Error

The error for point (x_i, y_i) and model (w, b) is:

$$e_i = y_i - (wx_i - b)$$

The L.S.E. or objective function is:

$$E(w,b) = \sum_i^N e_i^2$$

To optimize, we minimize E, i.e. taking the derivatives relative to tuning params.

Solve $\frac{dE}{db} = 0$ $\quad b^* = \hat{y} - w\hat{x}$, $\hat{y}, \hat{x}$ are average of y_i, x_i

Solve $\left.\frac{dE}{dw}\right|_{b^*} = 0$ $\quad w^* = \frac{\sum_i (y_i - \hat{y})(x_i - \hat{x})}{\sum_i (x_i - \hat{x})^2}$

Note: L.S.E. is very sensitive to outliers.

Multidimensional Input 1D Outputs

Now there are more tuning parameters.

We solve:

$$y = w_1 x_1 + w_2 x_2 + \dots + w_n x_n$$

$$= \begin{bmatrix} w_1 & w_2 & \dots & w_n & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ b \end{bmatrix}$$

For input $\vec{x}$, output y and parameters $\vec{w}$, b

The objective function:

$$E(\vec{w}) = \sum_i^N (y_i - (\vec{w}^T \vec{x}_i))^2$$

$$= \| \vec{y} - X\vec{w} \|^2$$

] X is matrix of $\vec{x}_i$

$$= (\vec{y} - X\vec{w})^T(\vec{y} - X\vec{w})$$

$$① \; = \vec{y}^T\vec{y} - 2\vec{y}X\vec{w} + \vec{w}^T X^T X\vec{w}$$

$\vec{y}$ is vector of y_i

w is vector of w_i

By applying : $\frac{d(a^T x)}{dx} = \frac{d(x^T a)}{d(x)} = a$

$$\frac{d(x^T A x)}{dx} = (A + A^T)x$$

to ①, we get optimal $\vec{w}^* = (X^T X)^{-1} X^T \vec{y}$

Multidimentional Input with Multidimensional Output

Thus solve :

$$\vec{y} = W^T \vec{x}$$

$$= \begin{bmatrix} | & | & & | \\ W^T x_1 & W^T x_2 & \cdots & W^T x_n \\ | & | & & | \end{bmatrix}$$

where $\vec{x}$ is output, $\vec{y}$ is input and W is matrix of tuning variables

The objective function:

$$E(W) = \sum_i^N \sum_j^k (y_{i,j} - (\vec{w}_j^T \vec{x}_i))^2$$

$$= \sum_j^k (\vec{y}_j - (X\vec{w}_j))^2$$

$$= \| Y - XW \|_F^2$$

$$= \sum_{i,j} (Y - XW)_{i,j}^2$$

Finally, the optimal solution is:

$$W^* = (X^T X)^{-1} X^T Y$$

[illegible]