TAM 598 Lecture 13:

Supervised Learning – Linear Regression via Least Squares

Announcements:

- HW 3 covers lectures 8-12 ; due on Mar 12
- One recorded class over spring break
- No class Monday Mar 24^{th}

supervised learning - given pairs of x and y observations, find the map between x and y.

classification
regression

I. Linear Regression via Least Squares - Supervised Learning

given n d-dimensional inputs $\underset{\sim}{x}_{1:n} = \{\underset{\sim}{x}_1, \underset{\sim}{x}_2, \ldots, \underset{\sim}{x}_n\}$

and n outputs : $y_{1:n} = \{y_1, \ldots, y_n\}$

"linear" means that the map from $\underset{\sim}{x}$ to y is linear

Two common ways to train a supervised learning model

(1) optimization

(2) probabilistic inference :

II. Linear Regression with a single variable - Example

$$L(\underset{\sim}{w}) = \sum_{i=1}^{n} \left(y_i - w_0 - w_1 x_i \right)^2 = \| \underset{\sim}{y} - \underset{\approx}{X} \underset{\sim}{w} \|^2$$

$$L = \| \underset{\sim}{r} \|^2 = \underset{\sim}{r}^T \underset{\sim}{r}$$

$\underset{\sim}{y}$ = vector of observations

$\underset{\sim}{w}$ = weight vector

$\underset{\approx}{X}$ = design matrix

Example 2: how things can go wrong

try instead: $y_i = -0.5 + 2x_i + 2x_i^2 + \epsilon_i$

and we still try to fit a linear model.

This is underfitting

III. Polynomial Regression

$$y_i = -0.5 + 2x_i + 2x_i^2 + \epsilon_i$$

Now we try $y = w_0 + w_1 x + w_2 x^2$ and minimize the square loss

Example 2 : what if we used a higher degree polynomial?

IV The Generalized Linear Model

eg) multi-dimensional polynomials

eg) radial basis functions

eg) fourier series

How to fit the generalized model using least squares:

quadratic loss function

$$\mathcal{L}(\underset{\sim}{w}) =$$

To minimize the loss:

1

IV. Measures of Predictive Accuracy

(1) Training and test datasets

- not a good idea to test how good your model is using the training dataset
- need to test model on unseen data - a test data set
- often, take your data and split (eg 70% train, 30% test)
- can plot predictions vs. observations

(2) Mean-squared error - scalar measure of goodness of fit

hard to understand the absolute meaning of MSE, so we can instead use relative MSE (RMSE)

$$\text{RMSE} = \frac{\text{MSE of your model}}{\text{MSE of simplest possible model}}$$

if $\text{RMSE} < 1 \rightarrow$

$\text{RMSE} > 1 \rightarrow$

(3) Coefficient of Determination R^2