ECE-GY 6143

Lecture 2

- [x] Homework logistics
- [x] Recap : Nearest neighbors
- [x] Regression.

Recap :

Data $\longrightarrow$ ML $\longrightarrow$ Actionable information

1) Data representation
2) Measure of goodness.
3) Algorithm to optimize for this measure.

Warm up : Nearest neighbor ~~s~~ search

Input: Dataset of documents.

Document $\longrightarrow$ vector of term frequencies

$$x_i(j) = tf_i(j)$$

"Term frequency".

$i \rightarrow i^{th}$ document

$j \rightarrow j^{th}$ word.

$$x_i(j) = tf_i(j) \cdot idf(j)$$

"Inverse document frequency".

$$idf(j) = \log\left(\frac{n}{n_j}\right)$$

TF-IDF representation of documents

Pros of NN

- Simple.
- Robust

Cons of NN

- $O(nd)$ time per test instance.

Regression

Dataset

Data → ← label.

(x_1, y_1)
(x_2, y_2)
(x_3, y_3)
$\vdots$
(x_n, y_n).

Goal of regression : Find a function f

$y_i \approx f(x_i)$.

Applications:

- Classification of image
  $x \to$ image, $y \to$ cat/no cat
- Retail pricing
  $x \to$ time
  $y \to$ price of product
- Weather
  $x \to$ location
  $y \to$ rainfall.
- Sentiment
  $x \to$ sentence
  $y \to$ +/- sentiment

Not well posed !!

Fix

Regression: find a function f that belongs to a class of functions $\mathcal{H}$

Hypothesis class

s.t. $y_i \approx f(x_i)$.

Example: ① $\mathcal{H}$: set of all linear functions.

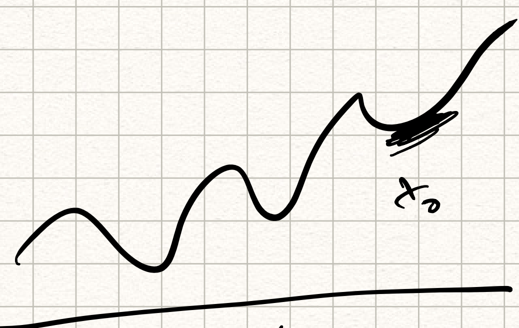
② $\mathcal{H}$: set of neural networks. of a particular architecture.

Linear models

$\mathcal{H}$: set of linear functions.

Why linear models?

- Simplicity.
- Stable behavior. [changes in input $\propto$ changes in output].
- Easy to compute.
- Interpretable.
- Building blocks of more complex phenomena.



$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \dots$$

Linear regression (univariate)

$x \rightarrow$ scalar
$y \rightarrow$ scalar.

(x_1, y_1)
(x_2, y_2)
$\vdots$
(x_n, y_n)

Step 1: Representation.

$$y = w_0 + w_1 x$$

Step 2: Measure of goodness.

e.g. MSE [mean squared error].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(w_0 + w_1 x_i - y)^2 \quad \Bigg\| \quad E_y(\hat{y} - y)^2$$

Step 3: Find best w_0, w_1 that minimizes mean square error.

Take partial derivative of MSE w.r.t. w_0 & w_1, equate to 0.

$$\frac{\partial MSE}{\partial w_0} = 0$$

2. $$\frac{1}{n}\sum_{i=1}^{n}(w_0 + w_1 x_i - y_i) = 0.$$

Solve for w_0.

$$w_0 = \frac{\sum_i y_i}{n} - w_1 \frac{\sum_i x_i}{n}.$$

$$w_0 = \bar{y} - w_1 \bar{x}$$

$$\frac{\partial MSE}{\partial w_1} = 0$$

$$2 \cdot \frac{1}{n} \sum_{i=1}^{n} x_i (w_0 + w_1 x_i - y_i) = 0$$

$$\frac{1}{n} \sum_i x_i (\bar{y} - w_1 \bar{x} + w_1 x_i - y_i) = 0$$

$$w_1 \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}) x_i = \frac{1}{n} \sum_i (y_i - \bar{y}) x_i$$

$$\boxed{w_1 = \frac{\overbrace{\frac{1}{n} \sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}}^{\text{Cross-Covariance of } x \,\&\, y.}}{\underbrace{\frac{1}{n} \sum_i x_i^2 - \bar{x}^2}_{\text{Variance of } x}}}$$

Variance

$$\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{1}{n} \sum_i (x_i^2 - 2 x_i \bar{x} + \bar{x}^2)$$

$$= \frac{1}{n} \sum_i x_i^2 - \bar{x}^2$$

Plug in values of w_0, w_1 in MSE

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - (w_0 + w_1 x)\right]^2$$

$\sigma_x^2 \rightarrow$ variance of x

$\sigma_y^2 \rightarrow$ variance of y.

$\sigma_{xy} \rightarrow$ cross covariance.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - (\bar{y} - w_1\bar{x} + w_1 x_i)\right]^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left[(y_i - \bar{y}) - \frac{\sigma_{xy}}{\sigma_x^2}(x_i - \bar{x})\right]^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sigma_{xy}}{\sigma_x^2}\right)^2 (x_i - \bar{x})^2$$

$$- 2\,\frac{1}{n}\sum_{i}\frac{\sigma_{xy}(y_i - \bar{y})(x_i - \bar{x})}{\sigma_x^2}$$

$$= \sigma_y^2 + \frac{\sigma_{xy}^2}{\sigma_x^4}\cdot\sigma_x^2$$

$$- 2\,\frac{\sigma_{xy}^2}{\sigma_x^2}$$

$$= \sigma_y^2 + \frac{\sigma_{xy}^2}{\sigma_x^2} - 2\frac{\sigma_{xy}^2}{\sigma_x^2}$$

$$MSE = \sigma_y^2 - \frac{\sigma_{xy}^2}{\sigma_x^2}$$

$$\boxed{\frac{MSE}{\sigma_y^2} = 1 - \frac{\sigma_{xy}^2}{\sigma_x^2\sigma_y^2}}$$

$$\frac{E(\hat{y}-y)^2}{var(y)}.$$

"Fraction of variance unexplained"

FVU

R^2 value.