Notations Overview

CSE446

Department of Computer Science & Engineering
University of Washington

January 2018

In Class

- For a **fixed** f (which does not depend on the training set D), the training error is an unbiased estimate of the expected error.
- Proof: Taking an expectation over the dataset D

$$
\begin{aligned}
\mathbb{E}_{\mathrm{D}}[\hat{\epsilon}(f)] &= \mathbb{E}[\frac{1}{N}\sum_n \ell(y_n, f(x_n))] && \text{Proof in CIML p15} \\
&= \frac{1}{N}\sum_n \mathbb{E}\ell(y_n, f(x_n)) && \text{Linearity of Expectation} \\
&= \frac{1}{N}\sum_n \epsilon(f) && \text{Showed in Class} \\
&= \epsilon(f)
\end{aligned}
$$

- Hard to read if you don't understand the notation!

Concepts and Notations

$$\mathbb{E}_{\mathrm{D}}[\hat{\epsilon}(f)] = \mathbb{E}[\frac{1}{N}\sum_n \ell(y_n, f(x_n))] = \frac{1}{N}\sum_n \mathbb{E}\ell(y_n, f(x_n)) = \frac{1}{N}\sum_n \epsilon(f) = \epsilon(f)$$

- Bias
  - The difference between this estimator's expected value and the true value of the parameter being estimated
  - Unbiased means expected value = true value
- Loss: $\ell(.,.)$
  - Function that quantifies the difference between the output and true values
  - In classification, it is equal to the number of misclassifications
- Estimator: $\hat{.}$
  - $\hat{y}$ is an estimator of the true y
  - $\hat{y}$ is an estimate based on known values, $\hat{y}$ = f(x)

Concepts and Notations

$$\mathbb{E}_{\mathrm{D}}[\hat{\epsilon}(f)] = \mathbb{E}[\frac{1}{N}\sum_n \ell(y_n, f(x_n))] = \frac{1}{N}\sum_n \mathbb{E}\ell(y_n, f(x_n)) = \frac{1}{N}\sum_n \epsilon(f) = \epsilon(f)$$

- Expectation: $\mathbb{E}_{\mathrm{D}}[.]$
  - '*Average*' over a distribution
  - Given a discrete random variable X.

$$\mathbb{E}[X] = \sum_x x Pr[X = x]$$

  - Example: Let $X1$ and $X2$ denote two independent rolls of a fair dice, what is the expected value of $X = X1 + X2$?

$$\mathbb{E}[X] = \sum_x x Pr[X = x] = 2\frac{1}{36} + 3\frac{1}{36} + ... + 12\frac{1}{36} = 7$$

  - Linearity of Expectation: $\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{n} Xi] = \sum_{i=1}^{n} \mathbb{E}[Xi]$

Another Proof

CIML page 15:

- Draw many pairs of (x, y) independently from dataset D, what would the expected loss be?

$$
\begin{aligned}
\mathbb{E}_{\mathrm{D}}[\ell(y, f(x))] &= \sum_{(x,y)\in D} \mathcal{D}(x, y)\ell(y, f(x)) && \text{Definition of Expectation} \\
&= \sum_{n=1}^{N} \mathcal{D}(x_n, y_n)\ell(y_n, f(x_n)) && \text{Discrete and finite } D \\
&= \sum_{n=1}^{N} \frac{1}{N}\ell(y_n, f(x_n)) && \text{Definition of } D \\
&= \frac{1}{N}\sum_{n=1}^{N} \ell(y_n.f(x_n)) && \text{Rearrange terms}
\end{aligned}
$$

- This is exactly the average loss on D.