CS 536 : Regression and Error

16:198:536

A Small Regression Example Consider regression in one dimension, with a data set $\{(x_i, y_i)\}_{i=1,\dots,m}$.

- Find a linear model that minimizes the training error, i.e., $\hat{w}$ and $\hat{b}$ to minimize

$$\sum_{i=1}^{m} (\hat{w}x_i + \hat{b} - y_i)^2. \tag{1}$$

- Assume there is some true linear model, such that $y_i = wx_i + b + \epsilon_i$, where noise variables ϵ_i are i.i.d. with $\epsilon_i \sim N(0, \sigma^2)$. Argue that the estimators are unbiased, i.e., $\mathbb{E}\left[\hat{w}\right] = w$ and $\mathbb{E}\left[\hat{b}\right] = b$. What are the variances of these estimators?

- Assume that each x value was sampled from some underlying distribution with expectation $\mathbb{E}[x]$ and variance $\text{Var}(x)$. Argue that in the limit, the error on $\hat{w}$ and $\hat{b}$ are approximately

$$\begin{aligned} \text{Var}(\hat{w}) &\approx \frac{\sigma^2}{m} \frac{1}{\text{Var}(x)} \\ \text{Var}(\hat{b}) &\approx \frac{\sigma^2}{m} \frac{\mathbb{E}\left[x^2\right]}{\text{Var}(x)}. \end{aligned} \tag{2}$$

- Argue that recentering the data ($x'_i = x_i - \mu$) and doing regression on the re-centered data produces the same error on $\hat{w}$ but *minimizes* the error on $\hat{b}$ when $\mu = \mathbb{E}[x]$ (which we approximate with the sample mean).

- Verify this numerically in the following way: Taking $m = 200, w = 1, b = 5, \sigma^2 = 0.1$.
  - Repeatedly perform the following numerical experiment: generate $x_1, \dots, x_m \sim \text{Unif}(100, 102)$, $y_i = wx_i + b + \epsilon_i$ (with ϵ_i as a normal, mean 0, variance σ^2), and $x'_i = x_i - 101$; compute $\hat{w}, \hat{b}$ based on the $\{(x_i, y_i)\}$ data, and $\hat{w}', \hat{b}'$ based on the $\{(x'_i, y_i)\}$ data.
  - Do this 1000 times, and estimate the expected value and variance of $\hat{w}, \hat{w}', \hat{b}, \hat{b}'$. Do these results make sense? Do these results agree with the above limiting expressions?

- *Intuitively, why is there no change in the estimate of the slope when the data is shifted?*

- Consider augmenting the data in the usual way, going from one dimensions to two dimensions, where the first coordinate of each $\underline{x}$ is just a constant 1. Argue that taking $\Sigma = X^{\text{T}}X$ in the usual way, we get in the limit that

$$\Sigma \to m \begin{bmatrix} 1 & \mathbb{E}[x] \\ \mathbb{E}[x] & \mathbb{E}[x^2] \end{bmatrix} \tag{3}$$

  Show that re-centering the data ($\Sigma' = (X')^{\text{T}}(X')$, taking $x'_i = x_i - \mu$), the condition number $\kappa(\Sigma')$ is minimized taking $\mu = \mathbb{E}[x]$.