Lecture 34, Apr 5, 2023

Linear Regression

- We have a set of data in the form of input-output pairs, $(x_i, y_i), i = 1, \ldots, n$; in general we want a function f such that $y = f(x)$ minimizes the errors $e_i = y_i - f(x_i)$
- For now we will talk about linear regression – assuming $y = ax + b$ so $e_i = y_i - (ax_i - b)$, so the total squared error is $\mathcal{E} = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2$
- Goal: find $\min_{a,b} \mathcal{E}$
  - $$\begin{aligned}\frac{\partial \mathcal{E}}{\partial a} &= \sum_{i=1}^{n} \frac{\partial}{\partial a}(y_i - ax_i - b)^2 \\ &= -\sum_{i=1}^{n} 2(y_i - ax_i - b)x_i \\ &= 0\end{aligned}$$
  - $$\begin{aligned}\frac{\partial \mathcal{E}}{\partial b} &= \sum_{i=1}^{n} \frac{\partial}{\partial b}(y_i - ax_i - b)^2 \\ &= -\sum_{i=1}^{n} 2(y_i - ax_i - b) \\ &= 0\end{aligned}$$
  - Rearrange and we get the normal equations: $\begin{cases} nb + a\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases}$
  - Since they are linearly independent we can directly solve; let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$
  - Solve: $\begin{cases} a = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \\ b = \bar{y} - a\bar{x} \end{cases}$
  - We know this is a minimum, since $\mathcal{E}$ is a simple convex quadratic
- With normal linear regression we only look at the vertical distances (i.e. errors in y); but we can improve it by looking at the normal (geometric) distance instead, which is called a *Deming regression*
  - In order to do this we also need to know the ratio of variances

Least Squares With Maximum Likelihood Estimation

- We assume each error e_i is a realization of a normal RV with mean 0 and variance σ^2
- $$\begin{aligned}L(e_1, \ldots, e_n; a, b) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{e^2}{2\sigma^2}} \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y_i - ax_i - b)^2}{2\sigma^2}}\end{aligned}$$
- If we maximize this, we get the same solution as the least squares approach
- We can think of this as assuming that each y_i is normally distributed, with mean $\mu = ax_i - b$ and uniform variance σ^2