Bais-Variance Tradeoff

Adil Khan

a.khan@innopolis.ru

Preliminaries - 1

- Let $D = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$ be our labelled dataset, where $\boldsymbol{x_i} \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$

- $D \sim P(\boldsymbol{x}, y)^N$

- For any given $\boldsymbol{x}$ there might not be a unique y – *<u>for example</u>*:

  - ➢Multiple houses of same size, number of rooms and neighborhood selling for different prices

Preliminaries - 2

- Thus for any given $\boldsymbol{x}$, we have a distribution over possible labels

- Therefore, let's define the following

$$\bar{y}(\boldsymbol{x}) = E_{y|\boldsymbol{x}}[y]$$

- Which represents the *expected/average* label that we would expect to obtain given $\boldsymbol{x}$

Preliminaries - 3

- Given our dataset D, we train a model f_D

- We can compute the generalization error of f_D as

$$E_{(\boldsymbol{x},y)\sim P}[(f_D(\boldsymbol{x}) - y)^2]$$

Preliminaries - 4

- Let's go back to our dataset D
- It is drawn from a distribution P^N and therefore it is a random variable
- If D is a random variable, then so is f_D
- Thus we can compute the *expected/average* model as follows

$$\bar{f}(\boldsymbol{x}) = E_{D\sim P^N}[f_D(\boldsymbol{x})]$$

Proof - 1

Add and subtract the expected classifier

$$E_{\boldsymbol{x},y,D}[(f_D(\boldsymbol{x}) - y)^2] = E_{\boldsymbol{x},y,D}\left[\left(f_D(x) - \bar{f}(\boldsymbol{x}) + \bar{f}(\boldsymbol{x}) - y\right)^2\right]$$

$$= E_{\boldsymbol{x},y,D}\left[\left(\left(f_D(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right) + \left(\bar{f}(\boldsymbol{x}) - y\right)\right)^2\right]$$

$$= E_{\boldsymbol{x},y,D}\left[\left(f_D(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right)^2 + \left(\bar{f}(\boldsymbol{x}) - y\right)^{\mathbf{2}} + 2\left(f_D(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right)\left(\bar{f}(\boldsymbol{x}) - y\right)\right]$$

$$= E_{\boldsymbol{x},D}\left[\left(f_D(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right)^2\right] + E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x}) - y\right)^{\mathbf{2}}\right] + 2E_{\boldsymbol{x},y,D}\left[\left(f_D(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right)\left(\bar{f}(\boldsymbol{x}) - y\right)\right]$$

Proof - 2

$$\boxed{E_{\boldsymbol{x},y,D}[(f_D(\boldsymbol{x})-y)^2] = E_{\boldsymbol{x},D}\left[\left(f_D(\boldsymbol{x})-\bar{f}(\boldsymbol{x})\right)^2\right] + E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x})-y\right)^{\mathbf{2}}\right] + \overline{2E_{\boldsymbol{x},y,D}\left[\left(f_D(\boldsymbol{x})-\bar{f}(\boldsymbol{x})\right)\left(\bar{f}(\boldsymbol{x})-y\right)\right]}}$$

Take the third term and solve it separately

$$E_{\boldsymbol{x},y,D}\left[\left(f_D(\boldsymbol{x})-\bar{f}(\boldsymbol{x})\right)\left(\bar{f}(\boldsymbol{x})-y\right)\right] = E_{\boldsymbol{x},y}\left[E_D\left[\left(f_D(\boldsymbol{x})-\bar{f}(\boldsymbol{x})\right)\right]\left(\bar{f}(\boldsymbol{x})-y\right)\right]$$

$$= E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x})-\bar{f}(\boldsymbol{x})\right)\left(\bar{f}(\boldsymbol{x})-y\right)\right] = 0$$

$$\boxed{E_{\boldsymbol{x},y,D}[(f_D(\boldsymbol{x})-y)^2] = E_{\boldsymbol{x},D}\left[\left(f_D(\boldsymbol{x})-\bar{f}(\boldsymbol{x})\right)^2\right] + E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x})-y\right)^{\mathbf{2}}\right]}$$

Proof - 3

$$\boxed{E_{\boldsymbol{x},y,D}[(f_D(\boldsymbol{x}) - y)^2] = E_{\boldsymbol{x},D}\left[\left(f_D(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right)^2\right] + \overline{E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x}) - y\right)^{\mathbf{2}}\right]}}$$

Take the second term and solve it separately

Add and subtract the expected/average label

$$E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x}) - y\right)^2\right] = E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x}) - \bar{y}(\boldsymbol{x}) + \bar{y}(\boldsymbol{x}) - y\right)^2\right]$$

Apply the same procedure as we did on slide 6, and the term "2ab" will yield 0, leaving us with

$$E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x}) - y\right)^2\right] = E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x}) - \bar{y}(\boldsymbol{x})\right)^2\right] + E_{\boldsymbol{x}}[(\bar{y}(\boldsymbol{x}) - y)^2]$$

Substituting this value in the top equations yields

$$\boxed{E_{\boldsymbol{x},y,D}[(f_D(\boldsymbol{x}) - y)^2] = E_{\boldsymbol{x},D}\left[\left(f_D(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right)^2\right] + E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x}) - \bar{y}(\boldsymbol{x})\right)^2\right] + E_{\boldsymbol{x}}[(\bar{y}(\boldsymbol{x}) - y)^2]}$$

Analysis - 1

$$E_{\boldsymbol{x},y,D}[(f_D(\boldsymbol{x}) - y)^2] = E_{\boldsymbol{x},D}\left[\left(f_D(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right)^2\right] + E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x}) - \bar{y}(\boldsymbol{x})\right)^2\right] + E_{\boldsymbol{x}}[(\bar{y}(\boldsymbol{x}) - y)^2]$$

Expected value of the squared distance between the model trained on D and the expected or average Model

Variance of the Model

Analysis – 2

$$E_{\boldsymbol{x},y,D}[(f_D(\boldsymbol{x})-y)^2] = E_{\boldsymbol{x},D}\left[\left(f_D(\boldsymbol{x})-\bar{f}(\boldsymbol{x})\right)^2\right] + \underline{E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x})-\bar{y}(\boldsymbol{x})\right)^2\right]} + E_{\boldsymbol{x}}[(\bar{y}(\boldsymbol{x})-y)^2]$$

Expected value of the squared distance between the prediction by the expected/average model and the expected/average label

Bias of the Model

Analysis - 3

$$E_{\boldsymbol{x},y,D}[(f_D(\boldsymbol{x}) - y)^2] = E_{\boldsymbol{x},D}\left[\left(f_D(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right)^2\right] + E_{\boldsymbol{x},y}\left[\left(\bar{f}(\boldsymbol{x}) - \bar{y}(\boldsymbol{x})\right)^2\right] + \underline{E_{\boldsymbol{x}}[(\bar{y}(\boldsymbol{x}) - y)^2]}$$

Expected value of the squared distance between the expected/average label and the given label

Noise in the Data