

0.1 De Schatter

omdat we meestal niet kunnen bepalen wat precies de variantie is van data hebben we nood aan een functie die de variantie schat, dit is de schatter. een veel gebruikte schatter is het steekproefgemiddelde. Deze wordt berekend met de formule:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

Waarbij \bar{X}_n de gemiddelde Stochastic, n het aantal datapunten, en X_i het i -de datapunt. Een schatter zonder bias is niet altijd beter.

0.2 accuraatheid

om de accuraatheid te bepalen van ons model splitsen we onze data in 3 delen namelijk: trainingsdata, validatiedata en testdata. vaak wordt gesplit in: 50% trainingsdata, 25% validatiedata, 25%. Nadat het model is getraind met de trainingsdata, wordt het gevalideerd via de validatiedata en eventueel opnieuw getraind als de MSE te groot is. De testdata is voor de gebruiker om te zien of het model werkt voor nog nooit geziene data

0.3 Bias

Bias kunnen we in woorden beschrijven als de gemiddelde fout op een voorspelde waarde. Bijvoorbeeld als de verwachte waarde 13 en de werkelijke waarde 65 zal de bias veel groter zijn dan wanneer de werkelijke waarde 14 is. Als we de bias wiskundig willen berekenen doen we dit met de volgende formule:

$$bias(\hat{\theta}) = E(\hat{\theta}) - \theta \quad (2)$$

waarbij θ gelijk is aan de variantie en $E(\hat{\theta})$ gelijk is aan de verwachte waarde van de schatter. Stel dat de bias nul is dan zal $\theta = E(\hat{\theta})$. Dan noemen we dit een onvertekende schatter (zie fig x,x+1). Een lage bias is niet altijd beter. Een lage bias kan lijden tot overfitting waarbij het model te veel is aangepast aan de trainingsdata en er dus slechte voorspellingen worden gedaan voor de validatiedata en de testdata. De bias kan ook worden gebruikt in de Formule voor de Mean-Squared Error(MSE) die bepaald hoe algemeen sterk het model is. De term Bias kan ook slaan op bepaalde stigmas die het model hanteert door niet representative testdata, volgens het principe garbage in, garbage out.

0.4 Mean-Squared Error

De MSE is een algemene maat voor hoe accuraat het model in het voorspellen van de klas tot waar een bepaalde test waarde tot behoort. De MSE kan op veel verschillende manieren worden berekend de meest courante is.

$$MSE(\hat{\theta}) = Var\hat{\theta} + (bias\hat{\theta})^2 \quad (3)$$

Waarbij $\hat{\theta}$ de verwachte waarde volgens de schatter is. Een lager MSE is steeds beter. als er geen bias is, is de variantie gelijk aan de MSE. Hier wordt dus bijna altijd naar gerefereerd om te bepalen welk model beter is.