Machine Learning @MIT Lecture Note 11 notes 1

Hien Nguyen (hienminhnguyen711@gmail.com)

September 30, 2020

Question 1.

Lecture note 11 Model selection criteria: Minimum description length and Feature selection.

Some note from Lecture note 11 @credit to MIT

Model selection criteria - compressed data - classification context - cost of error in training - cost of selected classifier

Simple two-part MDL P(yt|) =(yt,1)(1)(yt,1 - minimize encoding cost minlogP(yt)=logP(yt^)(4)||t=1t=1=l(y1, . . . , yn;^) - Find distribution in model -> minimize encoding length of data - Uniform distribution over possible discrete values DLofdatagivenmodelDLofmodel DL=l(y1, . . . , yn;^)+log(n+1

- description length of sequence

Universal coding, normalized maximum likelihood

- Find distribution in closest to just encoding the data with the best fitting distribution in model logPNML(y1, . . . , yn)minl(y1, . . . , yn;) =maxl(y1, . . . , yn;) - Distribution minimize maximum deviation -> normalized maximum likelihood distribution - minimax optimal coding length logPNML(y1, . . . , yn) =maxl(y1, . . . , yn;)+logexpmaxl(y1, . . . , yn;)(1

FEATURE SUBSET SELECTION - Feature vector -> useful for classification - deal with noisy and irrelevant feature data - weight how useful data - feature weighting problem - Kernel optimization - Naive bayes, assume none of feature and label independent - log likelihood - ML param estimate - Shannon entropy - Conditional entropy - log-likelihood if we assume feature to depend on label -> mutual infor

Mutual information