greatlearning

Machine Learning



End to End Machine Learning

- Models Covered
 - Naïve Bayes
 - QDA
 - Logistic Regression
 - Classification-RPART
 - Random Forest
 - Support Vector Machine
 - Gradient Boosting
 - XGB
 - KNN
- Basic Process
 - Use each of this method
 - Develop an evaluation matrix
 - Compare the models and choose the most appropriate

greatlearning

Practical Hands on

Dataset

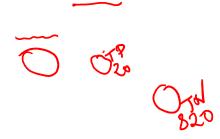
- Demo Data
- Socio Economic
 Parameters

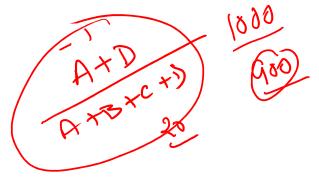
Columns

- Sex
- Age
- Education
- Unemployment Duration
- Income
- Marital Status
- Actively Involved in Sport
- Smoker?
- Alcohol Abuse
- Working Abroad
- Knowledge of English Language
- BMI

Describing Confusion Matrix

Reference		
Event	No Event	
A	В	
С	D	
	A	





The formulas used here are:

Sensitivity = A/(A+C)

Specificity = D/(B+D)

Prevalence = (A+C)/(A+B+C+D)

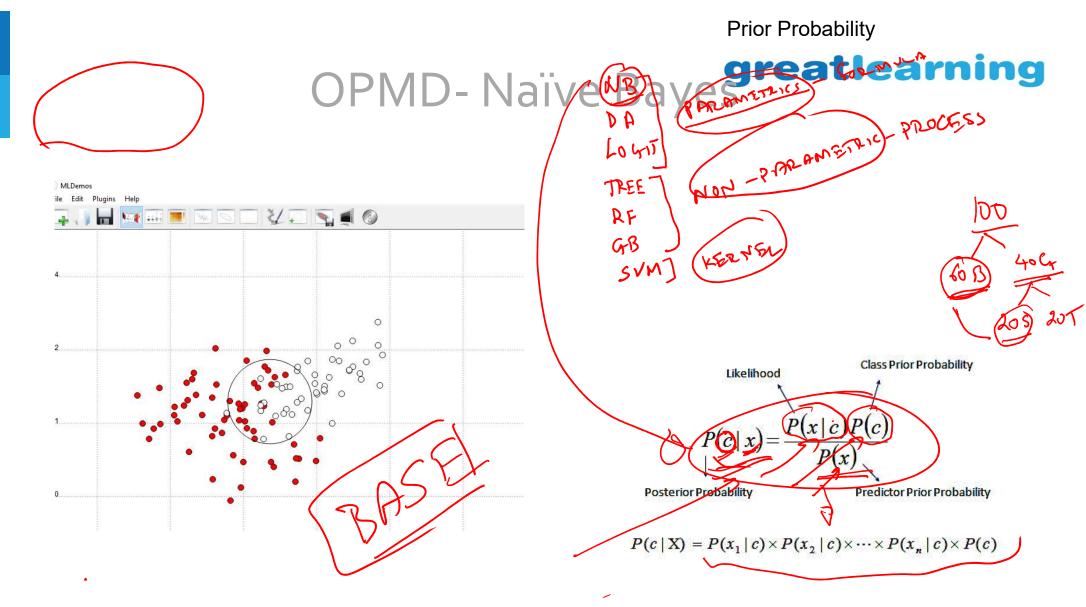
PPV = (sensitivity * Prevalence)/((sensitivity*Prevalence) + ((1-specificity)*(1-Prevalence)))

NPV = (specificity * (1-Prevalence))/(((1-sensitivity)*Prevalence) + ((specificity)*(1-Prevalence)))

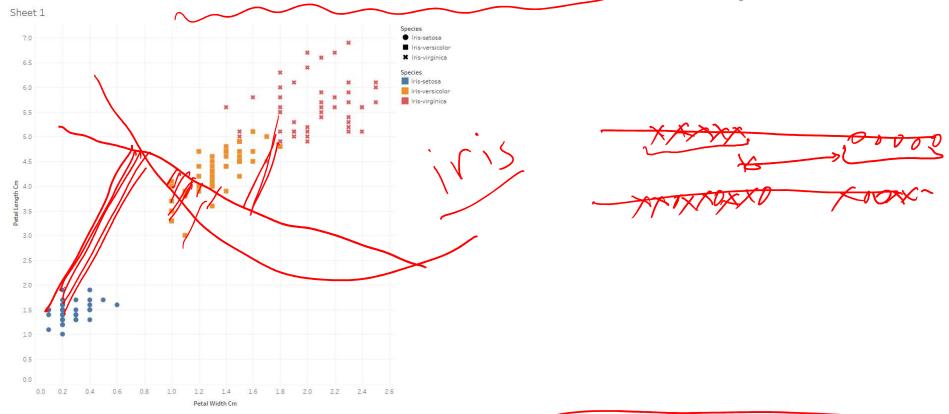
Detection Rate = A/(A+B+C+D)

Detection Prevalence = (A+B)/(A+B+C+D)

Balanced Accuracy = (Sensitivity+Specificity)/2



OPMD- Discriminant Analysis



Petal Width Cm vs. Petal Length Cm. Color shows details about Species. Shape shows details about Species

$$P(X|y=k)=rac{1}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \mathrm{exp}igg(-rac{1}{2}(X-\mu_k)^t\Sigma_k^{-1}(X-\mu_k)igg)$$
 All Rights Reserved. Unauthorized use or distribution prohibited.

Proprietary content. ©Great Learning. All Rights Reserved

OPMD-Logistic Regiession

2

6

