

Classification

| Classification Model | Pros | Cons |
|------------------------------|--|---|
| Logistic Regression | Probabilistic approach, gives informations about statistical significance of features | The Logistic Regression Assumptions |
| K-NN | Simple to understand, fast and efficient | Need to choose the number of neighbours k |
| SVM | Performant, not biased by outliers, not sensitive to overfitting | Not appropriate for non linear problems, not the best choice for large number of features |
| Kernel SVM | High performance on nonlinear problems, not biased by outliers, not sensitive to overfitting | Not the best choice for large number of features, more complex |
| Naive Bayes | Efficient, not biased by outliers, works on nonlinear problems, probabilistic approach | Based on the assumption that features have same statistical relevance |
| Decision Tree Classification | Interpretability, no need for feature scaling, works on both linear / nonlinear problems | Poor results on too small datasets, overfitting can easily occur |
| Random Forest Classification | Powerful and accurate, good performance on many problems, including non linear | No interpretability, overfitting can easily occur, need to choose the number of trees |