

Laboratório Aprendizagem de Máquina

Lab2: Impactos base de Aprendizagem

Para esse laboratório considere os seguintes classificadores:

- KNN
- Naïve Bayes
- Linear Discriminant Analysis
- Logistic Regression

Considere também as base de treinamento (20000 exemplos) e teste (58646 exemplos)

Compare o desempenho desses classificadores em função da disponibilidade de base de treinamento. Alimente os classificadores com blocos de 1000 exemplos e plote num gráfico o desempenho na base de testes e analise em qual ponto o tamanho da base de treinamento deixa de ser relevante.

Qual é o classificador que tem o melhor desempenho com poucos dados < 1000 exemplos.

Qual é o classificador que tem melhor desempenho com todos os dados.

Qual é o classificador é mais rápido para classificar os 58k exemplos de teste.

O que vc pode dizer a respeito das matrizes de confusão. Os erros são os mesmos para todos os classificadores quando todos eles utilizam toda a base de teste?