

# Impactos Base de Aprendizagem

Giovanni Rosa (GRR20148144)

Departamento de Informática, Universidade Federal do Paraná (UFPR)

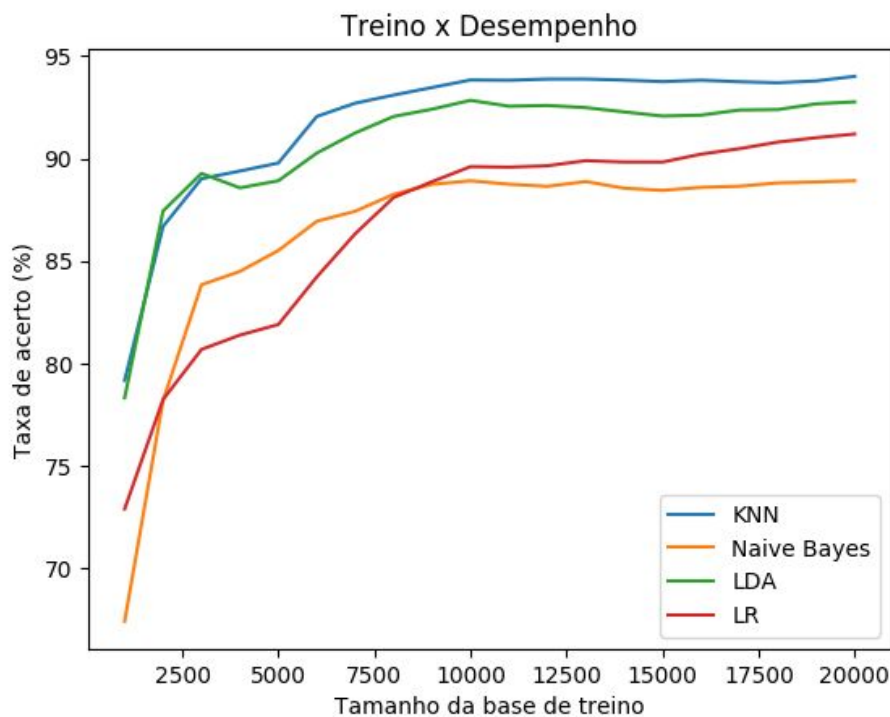
## 1.Introdução

Atividade da disciplina de Aprendizado de Máquina sobre representação KNN, segundo semestre de 2019.

## 2. Atividade

- Base de treinamento com 20.000 exemplos
- Base de teste com 58.646 exemplos
- kNN (k=3 e distância Euclidiana)
- Naive Bayes (Gaussian)
- Linear Discriminant Analysis (Least Squares)
- Logistic Regression (lbfgs e máximo de iterações = 1000)

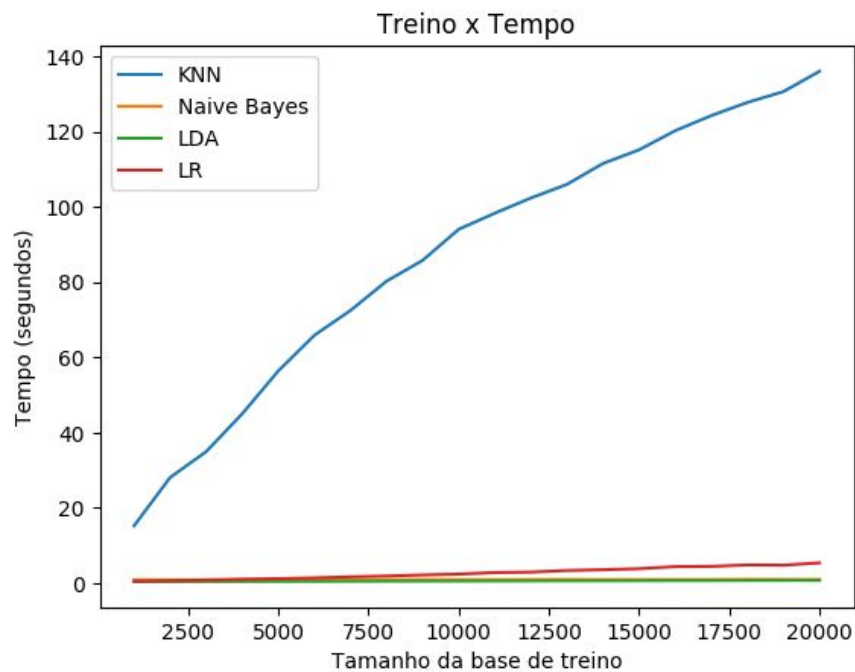
### 2.1. Comparação de Desempenho



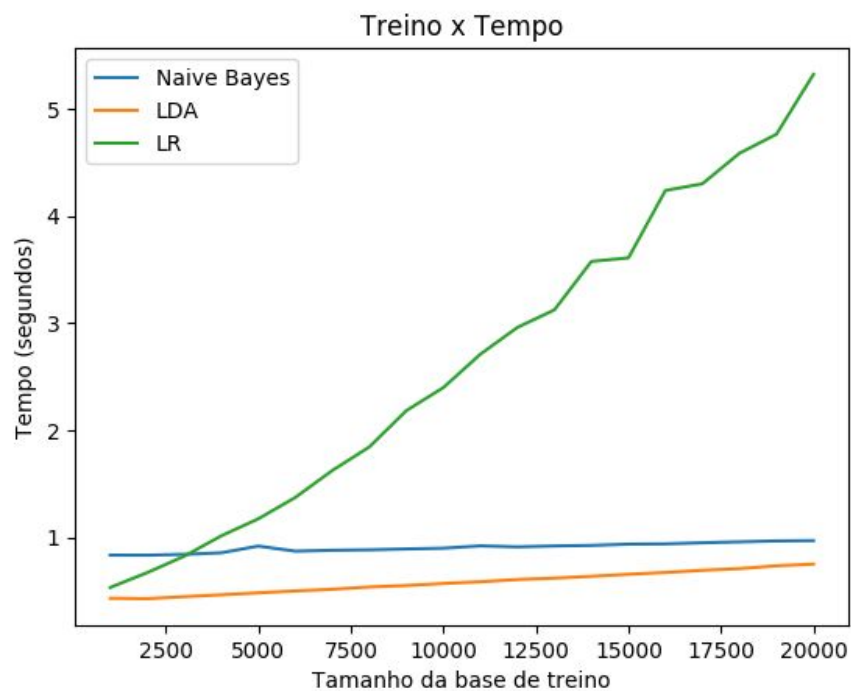
Podemos observar no gráfico que a partir de uma base de treinamento de 10000 registros, a taxa de acerto se torna estável, ou seja, a partir desse ponto o tamanho dessa base deixa de ser relevante.

Além disso, o classificador que tem o melhor desempenho tanto para poucos dados de treino (1.000 exemplos) quanto para a base completa (20.000) é o kNN, muito próximo do LDA.

## 2.2. Comparação de Tempo

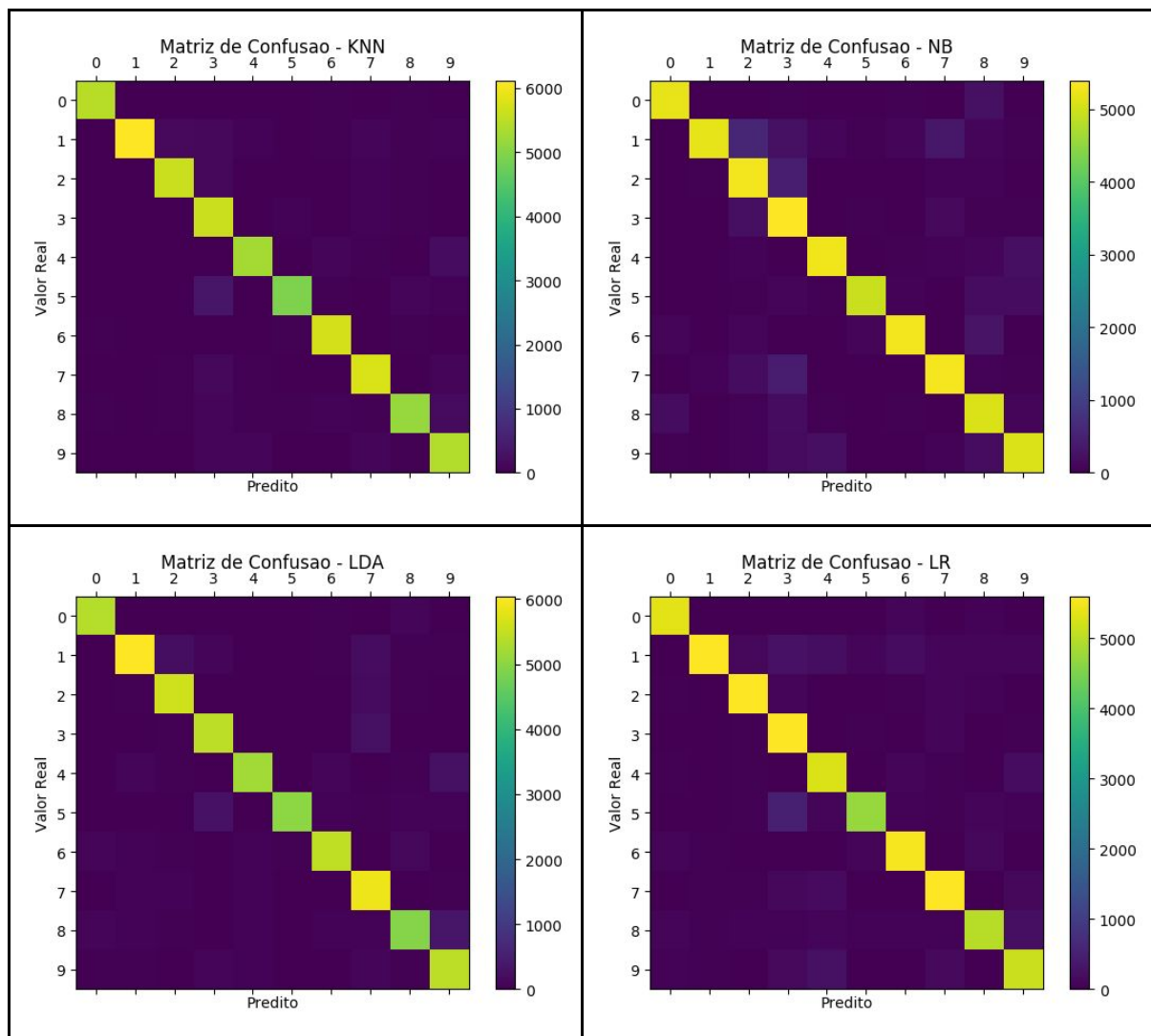


Nesse gráfico fica claro a disparidade do tempo gasto para ajuste e predição. O kNN leva quase 140 segundos para classificar a base completa (20.000 exemplos) enquanto os outros ficam entre 0,5 e 6 segundos.



Retirando o kNN do gráfico e aproximando aos outros, é possível concluir que o LDA é o classificador mais rápido tanto para poucos dados de treino (1.000 exemplos) quanto para a base completa (20.000 exemplos).

## 2.2. Comparação das Matrizes de Confusão



Analisando apenas as matrizes de confusão geradas para a base de treino completa (20.000 exemplos), notamos que a distribuição dos erros são diferentes entre os classificadores.

## 3. Conclusão

O classificador kNN apresenta a melhor taxa de acerto e o classificador LDA o melhor tempo. Unindo os dois dados, o LDA parece ser a melhor escolha por ter o melhor tempo e mesmo assim apresentar uma taxa de acerto próxima do kNN.

## 4. Bibliografia

<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

[https://scikit-learn.org/stable/modules/naive\\_bayes.html](https://scikit-learn.org/stable/modules/naive_bayes.html)

[https://scikit-learn.org/stable/modules/generated/sklearn.discriminant\\_analysis.LinearDiscriminantAnalysis.html](https://scikit-learn.org/stable/modules/generated/sklearn.discriminant_analysis.LinearDiscriminantAnalysis.html)

[https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html)