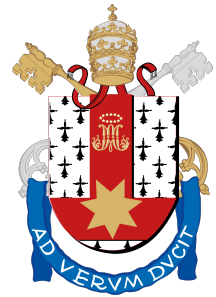


# Exercícios de Comitês

## Aprendizado de Máquina

Prof. Me. Otávio Parraga



## Exercício 1 - Comitês de Aprendizes

Responda as questões abaixo:

- (a) Por que *tree stumps* são utilizados em comitês de aprendizes baseados em Boosting?
- (b) Por que Árvores de Decisão são utilizadas em comitês de aprendizes baseados em Bagging?
- (c) Você treinou 5 classificadores utilizando algoritmos de aprendizagem diferentes, sendo que cada um desses classificadores possui acurácia próxima de 0.6. Dada essa situação, responda:
- (I) Supondo que os erros sejam altamente correlacionados, você espera melhoria na acurácia ao realizar um *ensemble* entre esses classificadores?
  - (II) Supondo que os erros sejam pouco correlacionados, você espera melhoria na acurácia ao realizar um *ensemble* entre esses classificadores?
- (d) Explique como introduzir diversidade em um comitê de aprendizes a partir de:
- (I) Manipulação do *dataset*.
  - (II) Manipulação de *features*.
  - (III) Manipulação do algoritmo de aprendizagem.
- (e) Qual a diferença do método de combinação de aprendizes por serialização e por votação?

---

## Exercício 2 - Bagging e Boosting

Identifique cada asserção abaixo como verdadeira ou falsa:

- (a) Bagging introduz diversidade no conjunto de dados por manipulação de *features*.
- (b) Ao usar Bagging por mais de uma iteração (*round*), mesmo que o aprendiz base utilizado tenha expressividade para aproximar apenas funções lineares, é possível aproximar uma

função não-linear.

**(c)** Em Bagging, é esperado que alguns dos aprendizes treinados tenham uma taxa de erro elevada devido ao peso aplicado em observações cujo erro dos aprendizes anteriores foi elevado.

**(d)** O processo de amostragem com reposição que é feito em cada rodada de Bagging se chama *bootstrapping*.

**(e)** *Random Forests* são diferentes do algoritmo de Bagging básico pois introduzem também uma manipulação das *features* do conjunto de dados para inserir variabilidade.

**(f)** *Out of Bag Error (OOB)* é computado em *Random Forests* utilizando apenas as observações que não foram consideradas na *bootstrap sample* que foi usada para treinar cada uma das árvores. OOB é um bom estimador para o erro em validação.

**(g)** O termo de importância, usualmente denotado por  $\alpha_i$ , do algoritmo de Boosting determina o peso do voto de cada um dos aprendizes treinados. Essa importância é inversamente proporcional à taxa de erro desse aprendiz.