# Dilema Viés-Variância

**Sobreajuste = menor Generalização**

**Caracterizado por Acuracidade onde? No conjunto de TREINAMENTO**

**Técnicas:**

**Separação de conjuntos de Treinamento e Teste**

**Cross Validation**

**Regularização – exemplo, R2 Ajustado**

# Respostas da prova...

**Responda com base nos conteúdos da aula**

# Entropia X Ganho de Informação

**Entropia = Quantidade de Informação, diversidade. Refere-se a um único atributo... E(S)**

**Ganho de Informação = O quanto a informação de um atributo contribuí para determinar outro. G(S,T)**

**Usado para a seleção de Atributos, ordem da Árvore de Decisão**

**Índice Gini, Error Classification formas equivalentes à Entropia para o cálculo de G(S,T).**

# Cross Validation

**N partições aleatórias...**

For i in range(nr_de_CV):

**N rodadas de treinamento,**

**N-1, partições para treinamento**

**1 partição de teste**

**Acuracidade... Média dos N treinamentos e Todos os Dados são ao menos uma vez empregados para teste.**

# Árvore de Decisão

**É um modelo partitivo e não utiliza os valores diretamente, mas as proporções ou probabilidade dos valores!**

**Ordem decrescente de Ganho de Informação dos nós**

**Entropia, Gini, Error Class = Hiperparâmetro!**

**Profundidade = Hiperparâmetro!**

**Como o K do K vizinhos, não há garantia de maior ou menor acuracidade**

# Random Forest

**É a média de várias diferentes árvores de decisão.**

**FLORESTA DE ÁRVORES ALEATÓRIAS**

**ERRO CUIDADO:** ~~ÁRVORES PARA PARTIÇÕES DOS DADOS~~

# Métricas de Eficiência dos Modelos

**Métricas de Classificação: Acuracidade, Precisão, Recall, F1-score essas e outras podem ser empregadas para avaliar ou selecionar um modelo!**

**Seleção de Hiperparâmetros!**

**Exemplos de Hiperparâmetros?**

**K do k-vizinhos e a função distância, max profundidade da Decision Tree, rede neural nr de neurônios e camadas, criterion (Decision Tree) etc.**

# GridSearch

**Seleção de Hiperparâmetros!**

- **varia K do Kvizinhos**
- **varia a profundidade da Árvore**

**Não Seleção de Modelos!!!**

- **troca de Kvizinhos**
- **para Árvore de Decisão**

# GridSearch

**For model in [‘Knn’, ‘Arvore de Decisão’, ‘Logistica’]:**

**gridSearchCV(model)**

**Knn, melhor k, melhor função distância**

**Árvore, melhor profundidade e critério**

**Logistica, melhor ...**

**De acordo com um critério que é uma métrica que pode ou não ser a acuracidade.**

# R

## Diferenças com relação ao Python...

**Não tem um padrão (padrão de estimadores do Scikit-learn) regra tão bem definida de como ‘programar’ os modelos pois existem muitos pacotes**

**Existem também diferenças das implementações e uso... Veja por exemplo as árvores de decisão!**

# Regressão Logística*(aula 8)

## Separador Linear (e Binário)

**P( x = ‘BENIGNO’ )**

**Mas o scikit-Learn faz Multiclasse com a regressão logística, COMO?**

**P( x = ‘setosa’ )**

**P( x = ‘virginica’ )**

**P( x = ‘setosa’ )**

# Classificação e Regressão

# Aprendizado Supervisionado