



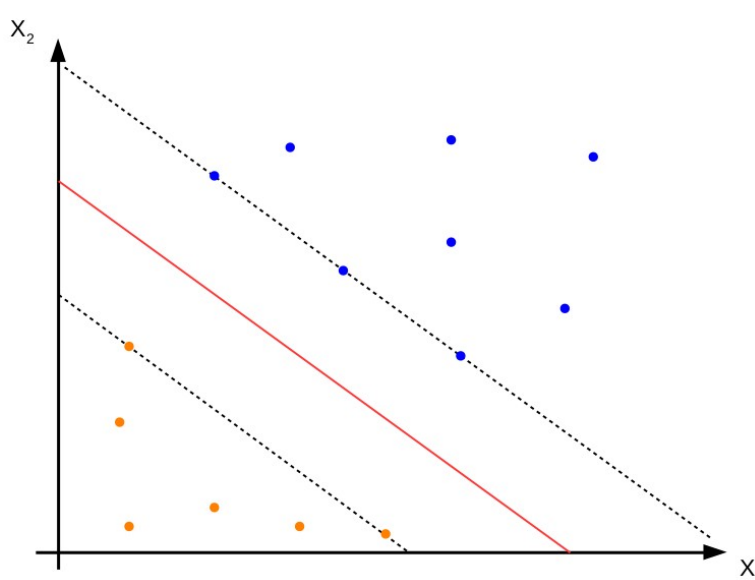
Lista de Exercícios 2

1. Defina aprendizado de máquina em termos de tarefas, experiências e métricas de desempenho. Exemplifique dois casos de aprendizado de máquina, explicitando a tarefa, a experiência de aprendizado e a métrica de desempenho.
2. Explique, com suas palavras, o que significam as expressões "linguagem de descrição dos dados" e "linguagem de descrição do modelo". Exemplifique.
3. O que é a capacidade de um modelo? Qual relação existe entre a capacidade do modelo e a complexidade de um problema de aprendizado?
4. Explique, com suas palavras, o que é *overfitting* e o que é *underfitting*. Que relação esses termos têm com a sua resposta para a questão anterior?
5. Baixe o conjunto de dados Iris, disponível no ColabWeb ou no repositório UCI.
 - a) Calcule o valor médio de cada atributo para as três classes (sugestão: abra o arquivo CSV em um editor de planilhas ou utilize agrupamentos do Pandas).
 - b) Calcule as variâncias de cada atributo.
 - c) Classifique o exemplo $\langle \text{sepal_length}=6,3; \text{sepal_width}=3,2; \text{petal_length}=4,0; \text{petal_width}=1,7 \rangle$ utilizando o modelo Naive Bayes. Assuma que atributos numéricos seguem uma distribuição gaussiana. Para estimar a probabilidade posterior de um atributo x possui o valor v para a classe C_k , utilize a seguinte equação:

$$p(x = v | C_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(v-\mu_k)^2}{2\sigma_k^2}}$$

Na qual v é o valor do exemplo para o atributo x , C_k é a k -ésima classe, μ_k é o valor médio do atributo x para exemplos da classe C_k e σ_k é o desvio padrão do atributo x para exemplos da classe C_k .

6. A figura na página seguinte ilustra um modelo de classificação SVM com separação linear para um problema de duas classes.
 - a) Identifique o vetor normal, o hiperplano, as margens de separação e os vetores de suporte.
 - b) Existem outros possíveis hiperplanos de separação para os dados da figura? Argumente.



- c) Caso sua resposta tenha sido afirmativa, explique por que o hiperplano selecionado é o mais adequado para os dados. Caso sua resposta tenha sido negativa, explique se existem condições alternativas que permitiriam a existência de um hiperplano de separação alternativo. Argumente.

7. Considere o seguinte conjunto de exemplos de treinamento

Instância	X_1	X_2	Classe
1	V	V	+
2	V	V	+
3	V	F	-
4	F	F	+
5	F	V	-
6	F	V	-

- a) Qual é a entropia dessa coleção com respeito à classe?
- b) Construa uma árvore de decisão considerando ganho de informação como métrica de desempenho do indutor.
- c) Gere regras de conhecimento utilizando o algoritmo de cobertura e considerando precisão como métrica de desempenho do indutor.
- d) Assumindo $V=1$ e $F=0$, você utilizaria um *kernel* polinomial para induzir um classificador SVM nesses dados? Por quê?