

Homework - GC - sobre - seleção - de - modelo

Q1)

Com $K=10$ folds, o conjunto de treinamento é dividido em 10 partes iguais e cada uma é usada como conjunto de validação em uma iteração.

$$\begin{array}{r} 720 \text{ (nº de hiperparâmetros)} \\ \times 10 \\ \hline 7200 \text{ modelos} \rightarrow 7200 \text{ horas} \end{array}$$

Ou seja, 300 dias é a duração total da tarefa

Q2)

Primeiramente dividiria os dados em três conjuntos; realizaria a busca de hiperparâmetros usando a validação cruzada com K -folds no conjunto de treino; escolheria os melhores hiperparâmetros com base no desempenho na validação cruzada e por fim, treinaria o modelo usando os melhores hiperparâmetros encontrados.

Q1)

a) Para cada configuração dos hiperparâmetros, serão treinados 6 modelos SVM, onde 6 é o número de folds da validação cruzada.

b) São 6 SVM, combinados com 3 "C" e 2 "gamma"

$$N^\circ \text{ de SVM} = 6 \times 3 \times 2 = 36$$

c)

$$\begin{aligned}\text{Tempo total} &= 36 \times 40 \text{ minutos} \\ &= 1440 \text{ minutos}\end{aligned}$$

d) Com um conjunto de validação seria necessário treinar um modelo SVM para cada combinação "C" e "gamma".

$$\begin{aligned}N^{\circ} \text{ de SVM} &= 2 \times 3 \\ &= 6 \text{ SVM}\end{aligned}$$

Q2)

Porque a metodologia usada é muito demandada, impraticável para conjunto de dados maiores, além de, alto custo computacional. O "leave-one-out" pode levar a resultados enviesados, pois só há um exemplo no conjunto de validação.

Uma alternativa é usar validação cruzada com K-fold, que é uma metodologia mais eficiente e que permite uma estimativa mais robusta de generalização.

Q7)

O underfitting ocorre quando o modelo não é suficientemente complexo para capturar a relação entre as variáveis de entrada e saída, resultando em baixa precisão. Um exemplo é a figura a).

O overfitting ocorre quando o modelo é muito complexo e ajusta-se excessivamente aos dados, levando a regiões de decisão muito complexas e com muitos detalhes, levando a um modelo com baixa capacidade de generalização. A figura c) se encaixa no overfitting.

O modelo b) é o que melhor apresenta as regiões de decisão.