

Conceitos:

- **Peritos:** Modelos especializados que são treinados para desempenhar tarefas e resolver problemas específicos, geralmente dentro de um domínio particular.
 - **Comités de peritos:** são combinações de vários modelos (peritos) para a criação de modelos mais poderosos. criam-se quando nenhum modelo/algoritmo apresenta o melhor desempenho para qualquer conjunto. esta combinação pode ter os peritos a trabalhar em paralelo - múltiplos peritos (em que os peritos trabalham todos ao mesmo tempo sobre os mesmos dados, e um módulo reúne as respostas e apresenta a decisão final) ou em série - múltiplas etapas (Cada perito dá pesos diferentes aos exemplos ou "foca-se" num sub-conjunto dos exemplos)
 - **Votação:** Simples (Todos os peritos têm o mesmo peso) ou Pesada (utilizadas probabilidades á posteriori).
 - **Bootstrap:** Dado um conjunto de dados de tamanho N, retiram-se aleatoriamente com reposição, N exemplos
 - **Estratégias:** métodos diferentes empregados na comniação de peritos para a formação de comités.
 - **BackPropagation:** Processo de atualização de pesos nas redes neuronais.
-

Estratégias de combinação de peritos:

- **bagging:** Bootstrap Aggregating. Múltiplos peritos, Cria L conjuntos de treino (método bootstrap), votação simples entre os L peritos. Bom para algoritmos instáveis (pequenas alterações nos dados podem provocar grandes diferenças no modelo). Robusto para dados com ruído.
 - **stacking:** Múltiplos peritos. Assume existencia de L peritos. Combinação entre peritos é aprendida e o combinador é treinado com um conjunto de dados não utilizado na construção dos peritos. Os peritos devem produzir previsões diferentes.
 - **boosting:** Múltiplas etapas. O comité é treinado com base nos erros dos classificadores anteriores. Votação pesada. Utiliza algoritmos de aprendizagem simples. Muito mau para ruído e outliers. Geralmente produz melhores classificadores que o bagging, pode sofrer de sobreajustamento.
 - **cascading:** Múltiplas Etapas. Peritos base ordenados por complexidade ou custo da representação. O perito P_j só é usado se os peritos precedentes não forem confiáveis. P_j tem uma confiança aliciada W_j , e é confiável se $W_j > 0_j$
-

RESOLUÇÃO DA SEGUNDA FREQ. MODELO 2023

1. Indique as principais diferenças entre bagging e boosting.

R: A estratégia bagging: tem combinação de Múltiplos peritos (os modelos trabalham em paralelo sobre os mesmos exemplos), Cria L conjuntos de treino através do método bootstrap (sendo L o numero de peritos), emprega votação simples na agregação dos peritos e é robusto para dados com ruído e bom para algoritmos instáveis.

Já a estratégia boosting: emprega combinação de Múltiplas etapas (os modelos executam passos diferentes e/ou trabalham sobre partes diferentes dos dados totais), O comité é treinado com base nos erros dos peritos anteriores, emprega votação pesada e é muito suscetível para dados com ruído e outliers.

Também é pertinente acrescentar que a estratégia de boosting, embora esteja mais suscetível ao sobreajustamento, geralmente cria melhores classificadores do que a estratégia de bagging

Em suma, as principais diferenças entre bagging e boosting são a sua lógica de combinação, o tipo de votação na agregação dos peritos, os tipos de dados e modelos para o qual são mais próprios e o desempenho na criação de classificadores.

2. Considere uma rede neuronal usando um perceptron base. Para a componente de limiar indique quais das seguintes funções $y=f(x)$ podem ser usadas:

- $y=ax+b$
- $y=1/(1+e^{(-x)})$ (sigmóide)
- $y= \tanh()$ (tangente hiperbólica)
- $y=ax^2$

Se alguma delas não for possível usar justifique o motivo.

R: Das funções demonstradas, a primeira, $y= ax+b$, é linear, e como é necessária uma função não-linear para a componente limiar, esta não poderá ser utilizada. Todas as outras funções são não-lineares, logo, podem ser utilizadas para a componente de limiar. note-se que a função $y =ax+b$, embora não linear, dependendo do problema, pode ser combinada com outras funções e camadas para gerar um output não linear, mas por si só não pode ser utilizada na componente limiar

3. Indique a resposta correta. “Nas redes neuronais a finalidade da backpropagation é”

- Evitar o sobre-ajustamento (backpropagation por si só não o faz)
- Aumentar a velocidade de convergência (embora ajude, não é o seu objetivo principal)
- Otimização dos pesos da rede (slides e chat gpt)
- Selecionar o subconjunto de teste (backpropagation não tem nada a ver com isso. selecionar o subconjunto de teste até é feito antes de poder ser aplicado o backpropagation, pois é geralmente feito antes do treino)

R: Otimização dos pesos da rede.

4. R: Na da esquerda existe sobre ajustamento, pois as previsões do algoritmo estão claramente demasiado ajustadas aos dados de treino, não correspondendo corretamente às previsões esperadas para os dados de teste aos dados de teste.

5. Considere os seguintes casos, e responda às questões para cada um deles.

- a) **2 algoritmos treinados num conjunto de dados de treino, foram ambos testados com conjunto de validação; uma vez escolhido o melhor deles, foi re-treinado com conjunto de treino+teste com 2 parâmetros diferentes, e foram testados finalmente com o conjunto de teste. Por fim, escolheu-se o melhor deles, e apresentou-se como resultado final. O procedimento é correto? Será que existe drenagem de informação do conjunto de teste? O resultado poderá ser enviesado (biased)?**

R: Não, o procedimento não é correto e não está de acordo com a boa prática em termos de drenagem de informação. Existe drenagem porque no segundo treino, um dos modelos foi treinado com o conjunto de dados treino+teste, então há drenagem de informação sobre o conjunto de teste para esse modelo, o que leva ao sobreajustamento do modelo aos dados de teste e, conseqüentemente, ao enviesamento da escolha pelo desempenho otimista do modelo quanto aos dados de teste, sendo o resultado final biased, mais virado para o modelo que foi treinado com o conjunto treino+teste.

- b) **1 algoritmo foi testado com validação cruzada (10 folds) num conjunto de dados, repetiu-se o procedimento com 4 valores diferentes do parâmetro X de regularização, selecionou-se o melhor deles, e apresentou-se o valor de desempenho. procedimento é correto? Será que existe drenagem de informação do conjunto de teste? O resultado poderá ser enviesado (biased)?**

R: Não só o procedimento é o correto (na generalidade dos casos), como a validação cruzada ajuda a mitigar a drenagem e o bias, pois “fabrica” vários conjuntos de teste com os mesmos dados. Desde que os modelos não sejam treinados com os dados com os quais vão ser testados, então a drenagem e o viés, neste caso, não é um problema

- c) 4 algoritmos foram treinados e testados com validação cruzada (10 folds) num conjunto de dados, selecionou-se o melhor deles, e apresentou-se o valor de desempenho. O procedimento é correto? Será que existe drenagem de informação do conjunto de teste? O resultado poderá ser enviesado (biased)?

R: o procedimento está correto. Não há drenagem desde que se mantenha separação entre os dados de teste e de treino

6. Indique qual das definições corresponde à macro média de F1 para avaliação de desempenho multiclasse:

- média não pesada de F1 de cada classe (slides e chatgpt)
- média de F1 pesada pelo nº de exemplos de cada classe (real) (nunca se há de medir desempenho com resultados reais)
- média de F1 pesada pelo nº de exemplos de cada classe (no resultado do classificador) (isso é macromédia pesada)
- cálculo de F1 depois de somar o nº total de verdadeiros positivos; e para falsos positivos, e falsos negativos somar o nº total de casos mal classificados, de todas as classes. (nem sei, mas não é)
- média não ponderada entre F0 e F2 (questão merda meu? tb não é)

8. Considere um problema de classificação com 2 classes. Uma com 100 instâncias e outra com 9900 instâncias. Obteve um classificador com 99% de exatidão. Indique se cada uma das afirmações é Verdadeira ou Falsa.

- a) O desempenho do sistema é claramente muito bom.

R: Falso - A exatidão como medida de desempenho do classificador pode ser enganadora, já que se o classificador escolhesse, para todos os exemplos, a classe majoritária, obteria na mesma uma exatidão de 99% pois escolheria certo para 9900 instâncias e errado para 100.

- b) Não tenho informação para saber se o desempenho é bom.

R: Verdadeiro - A exatidão por si só pode não ser a melhor indicação de desempenho, vai em conta ao que escrevemos acima.

- c) Se usasse um classificador aleatório o desempenho seria muito inferior.

R: Caso a métrica de desempenho seja a exatidão, então Verdadeiro - ao usar um classificador aleatório obteríamos uma exatidão de aproximadamente 50%, que é bem inferior a 99%.

- d) Se optasse por escolher a classe majoritária esperaria uma exatidão de 50%.

R: Falso - ver resposta à b) : se escolhermos a classe majoritária obteríamos uma exatidão de 99%.