

### Questão 1

Uma questão determinante na qualidade de modelos de classificação é o quanto a base de treinamento é balanceada, ou seja, se o espaço de ocorrência das entidades a serem classificadas é adequadamente amostrado. Considerando as árvores de decisão, discuta o quanto ela é robusta a esse tipo de problema. Se ela for robusta, justifique a sua resposta, caso contrário, indique como o problema pode ser minimizado.

**R:** Árvores de decisões são efetivas e robustas quando a base de dados é balanceada. Em uma base de dados balanceada, os pontos de separação da árvore são escolhidos de maneira ótima para melhor realizar a separação em dois grupos com o mínimo de mistura.

### Questão 2

Modelos de classificação são severamente afetados pela ocorrência de paradoxos na base de dados de treinamento, ou seja, os mesmos atributos levam a classes diferentes. Considerando a técnica "stacking", discuta se, e como, essas técnicas minimizam os problemas causados por paradoxos, ou explique porque não tratam.

**R:** A técnica de stacking realiza a combinação das previsões de dois ou mais algoritmos de classificação. As vantagens de realizar o "stacking" é o aproveitamento das capacidades de uma variedade de algoritmos de bom desempenho de classificação, o que pode resultar em previsões melhores do que quando previstas em apenas um único modelo – podendo minimizar os problemas de paradoxos.

### Questão 3

Randomização é uma estratégia popular em mineração de dados, em geral permitindo reduzir o viés de modelos de classificação e regressão. Considerando dois casos em particular, as árvores de decisão aleatórias e dropout em redes neurais profundas, explique como a randomização associada a cada uma delas auxilia em obter modelos melhores.

**R:** Dependendo das aplicações, as árvores de decisões ficam extensas e cada vez mais profundas, o que pode ser um problema: quanto mais profunda for a árvore, mais suscetível ao *overfitting* ela é. Desta forma, os algoritmos de árvores de decisões aleatórias como o Random Forest, por exemplo, utilizam a randomização de forma a reduzir o *overfitting*, além de uma maior simplicidade e acurácia, tornando o modelo

mais confiável e melhor. De forma semelhante, o *dropout* em uma rede neural é geralmente utilizado em cada camada da rede e possui o objetivo de reduzir o *overfitting* realizando *drops* em unidades aleatórias das camadas. A utilização da técnica de *dropout* força a rede neural a aprender características mais robustas.

#### Questão 4

A utilização de kernels no contexto de técnicas baseadas em álgebra linear como o SVM é uma estratégia consagrada. Apesar do kernel utilizado, as características do SVM como tolerância a ruídos se mantêm. Discuta como exceções e ruídos podem ser tratados no contexto de SVM+kernels e como a calibração dos algoritmos (por exemplo a constante C) pode levar em consideração o kernel utilizado.

**R:** A constante C informa à otimização do SVM o quanto você deseja evitar a classificação incorreta de cada exemplo de treinamento. Para grandes valores de C, a otimização escolherá um hiperplano de margem menor se esse hiperplano fizer um trabalho melhor em obter todos os pontos de treinamento classificados corretamente. Por outro lado, um valor muito pequeno de C fará com que o otimizador procure um hiperplano de separação de margem maior, mesmo que esse hiperplano classifique incorretamente mais pontos. Para valores muito pequenos de C, você deve obter exemplos classificados incorretamente, geralmente mesmo se seus dados de treinamento forem linearmente separáveis. Desta forma, a constante C é o principal fator para que a reta do hiperplano seja traçada de forma a tratar as exceções e ruídos presentes na base de dados.

#### Questão 5

A utilização de redes neurais de múltiplos níveis foi limitada por muito tempo em razão da impossibilidade de controlar a magnitude do gradiente, tanto do ponto de vista de crescer quanto reduzir muito rápido, que é conhecido como o problema do "vanishing/exploding gradient". Explique o que é esse problema em termos da impossibilidade desse controle e descreva as estratégias que foram adotadas para minimizar o problema e viabilizar as redes neurais profundas.

**R:** Quando uma rede neural profunda é treinada com aprendizagem baseada em gradiente e *backpropagation*, a regra da cadeia é utilizada para que as camadas mais profundas da rede passem por multiplicações de matrizes contínuas para calcular suas

derivadas. Desta forma, em uma rede neural de  $n$  camadas ocultas, as  $n$  derivadas serão multiplicadas juntas. Se as derivadas forem grandes, o gradiente aumentará exponencialmente conforme propagamos para baixo no modelo até que eles explodam (exploding gradient problem). Alternativamente, se as derivadas forem pequenas, então o gradiente diminuirá exponencialmente conforme propagam através do modelo até que eventualmente desapareça (vanishing gradient problem). Para a minimização desse problema, as estratégias que foram adotadas foram: redução do número de camadas, inicialização/regularização dos pesos e o recorte do gradiente (para o caso da explosão).

## Questão 6

Regularização é uma estratégia popular de simplificar modelos e aumentar a sua robustez. Durante o curso vimos regularizações L1 e L2 em mais de um cenário. Suponha que você tem que avaliar, qualitativa e quantitativamente, a utilização de uma regularização L3, ou seja, a função objetivo da regressão, linear ou logística, contempla a norma elevado ao cubo. Discuta a efetividade e os resultados da regularização L3, em contraponto às regularizações L1 e L2.

**R:** As regularizações L1 e L2 se diferem no que chamamos de penalização. Enquanto que a regularização L1 adiciona um “valor absoluto de magnitude” de coeficiente como penalização para a função de perda, o algoritmo L2 adiciona um “valor quadrático de magnitude” de coeficiente para a função de perda. De forma semelhante, o algoritmo L3 adicionaria um valor elevado ao cubo a função de perda como penalização mas isso tornaria a função muito complexa, obtendo resultados mais complexos ainda. Além de não ser muito efetiva e útil, visto que as regularizações L1 e L2 servem para a grande maioria dos problemas.