Materiais:

Q: Onde encontro os materiais de apoio e os arquivos notebooks para execução dos exercícios?

R: O material de apoio consta no ícone de folha a4 ao lado do título da disciplina.

Ferramentas:

Conceitos e Exercícios:

Q: No caso do alfa da função de custo (15 minutos de vídeo), por exemplo, que ferramenta poderia ser usada para calculá-lo? É possível obter esse dado usando Sklearn, por exemplo?

R: A taxa de aprendizado (alfa) regula a intensidade das atualizações de pesos durante o processo de otimização. Trata-se de um hiper parâmetro, ou seja, não é aprendido durante o treinamento, e deve ser informado pelo usuário antes do processo iniciar. Geralmente utiliza-se valores pequenos, entre 0.1 e 0.00001. O valor apropriado para a taxa de aprendizado depende de diversos fatores, como: qual técnica de aprendizado será utilizada (regressão linear/logística, rede neural etc.), a complexidade do modelo (ex.: profundidade de uma rede neural), função de custo utilizada, entre outros. Para se obter uma estimativa da qualidade deste valor de hiper parâmetro, podemos utilizar algum protocolo e medida de avaliação (aula 6, parte 3 e 4) para testar diversos valores e encontrar valores apropriados. Falando especificamente no Scikit-Learn, recomendo a leitura deste material para entender as funções disponíveis: https://scikit-learn.org/stable/model_selection.html.

Q: A fórmula apresentada na aula 2 parte 3 no instante 37:51 do vídeo contém o elemento "-1" no denominador? A fórmula certa é realmente (rank - 1) / (Qtd.Valores - 1)?

R: Refazer o exemplo do slide para que a explicação fique mais clara.

Categorias = {pequeno, médio, grande}

Fórmula: (rank - 1) / (num_valores - 1)

Assumindo que os ranks são 1, 2 e 3 para pequeno, médio e grande, respectivamente:

- Pequeno: (1 - 1) / (3 - 1) = 0 / 2 = 0
- Médio: (2 - 1) / (3 - 1) = 1 / 2 = 0.5
- Grande: (3 - 1) / (3 - 1) = 2 / 2 = 1

Q: Na aula 2, parte 4, no instante 26:35 do vídeo o professor fala que a coluna " *Weighted distance (1/d)* " é o resultado do valor na coluna ** _Distance L2_ * *(1/("Distance L2"))* , desculpe se estou deixando passar algo, mas essa conta não resulta em 1? Aparentemente o que consta na coluna *"Weighted distance (1/d)"* é apenas a conta *(1/(Distance L2))*.

Para esse método de *voto ponderado* seria apenas *(1/(Distance L2))* e somar os resultados, tendo como vencedor o de maior pontuação?

R: Sim, sua interpretação da coluna "*Weighted Distance*" está correta, ela reflete somente o resultado do inverso da distância.

Agora sobre o cálculo de voto ponderado, você também está correto: bastaria somarmos os resultados dos inversos das distâncias para cada classe e verificar qual das classes possui a maior pontuação. Contudo, este resultado não representa uma probabilidade, pois não estaria normalizado. Para obter uma interpretação probabilística como saída, bastaria dividir a soma de cada uma das classes pela soma total de todos os inversos das distâncias.

Pegando como exemplo a tabela do slide 62:

- Benign: 2 (0.5)
- Malignant: 2.2360 (0.448)
- Malignant: 2.8284 (0.354)

Soma total dos inversos das distâncias: 0.5 + 0.448 + 0.354 = 1.302

Valor para a classe Benign: 0.5 / 1.302 = 0.3840 (38.40%)

Valor para a classe Malignant: (0.448 + 0.354) / 1.302 = 0.6159 (61.59%)

Q: Tempo 28:03. Achei que seria escolhido a menor distância (mais similar) e não a maior distância (menos similar), o certo seria definir como Benigno como está no slide não é mesmo?

Só seria escolhido a classe com a menor distância para o caso k=1. Para os demais casos (k>1), devemos contar o número de votos para cada classe, e escolher a classe com o maior número de votos. No caso do voto ponderado, bastaria somarmos os resultados dos inversos das distâncias para cada classe e verificar qual das classes possui a maior pontuação. Neste caso, o maior valor é o da classe Malignant (0.802). Contudo, este resultado não representa uma probabilidade, pois não estaria normalizado. Para obter uma interpretação probabilística como saída, bastaria dividir a soma de cada uma das classes pela soma total de todos os inversos das distâncias.

Pegando como exemplo a tabela do slide 62:

- Benign: 2 (0.5)
- Malignant: 2.2360 (0.448)
- Malignant: 2.8284 (0.354)

Soma total dos inversos das distâncias: 0.5 + 0.448 + 0.354 = 1.302

Valor para a classe Benign: 0.5 / 1.302 = 0.3840 (38.40%)

Valor para a classe Malignant: (0.448 + 0.354) / 1.302 = 0.6159 (61.59%)

Q: Professor, eu sempre posso utilizar a matriz de confusão como acurácia? para qualquer caso?imagino que nos casos com poucos dados, mas é sempre confiável usá-la?

R: A matriz de confusão em si não é uma métrica, mas sim um mecanismo usado para extrair informações sobre as saídas do modelo e extrair métricas de desempenho a partir disto. Como discutimos em aula, as métricas que devem ser usadas dependem de uma análise do contexto do problema. Por exemplo, se as classes do problema estão balanceadas, seria viável utilizar a acurácia. Mas é possível extrair outras medidas à partir da matriz de confusão, como precisão, revocação, etc.

**FAQ gerado com base em comentários até o dia 02/03/2022.*