

Mural de quests da Biblioteca Real (Quests 3)

Prof. Daniel R. Cassar

Além da resolução dos problemas, toda entrega deve ter o nome dos membros do grupo e pelo menos uma breve introdução, uma breve discussão e uma breve conclusão.

Data limite de entrega: 22:00 de 28 de outubro de 2024

1 As hipóteses reais

ID: Modelos 4

Proponente: Milu Iluminarius MLVI

Lore: Milu Iluminarius MLVI é o rei do reino de Lumi. Ultimamente, está bastante preocupado pois Vostrasirus Baba, o vidente mais famoso do reinado, previu o ataque iminente de um dragão! Um perigo!! Diz a lenda que Milu Iluminarius I, o primeiro de seu nome, virou rei pois foi o único que conseguiu remover A Grande Espada da Luz da Lua que estava fincada em uma pedra. Essa história na verdade é uma grande balela e ele apenas virou rei pois era detentor de terras.

Objetivo: Vossa Iluminação levantou as hipóteses abaixo durante sua última reunião com a cúpula de Ciência, Tecnologia e Encantamentos do reinado. Sua tarefa é responder ao menos uma das hipóteses formuladas pelo rei. Para esta quest, utilize o dataset que pretende usar no seu projeto de conclusão do semestre desta disciplina e escolha o algoritmo de aprendizado de máquina que preferir para os itens 1, 2 e 3.

1. Realizar normalização padrão dos atributos melhora a performance de modelos induzidos pelo algoritmo escolhido
2. Reduzir a multicolinearidade dos atributos melhora a performance de modelos induzidos pelo algoritmo escolhido (permitido escolher esta apenas caso haja multicolinearidade nos dados)
3. Reduzir a dimensionalidade do problema utilizando PCA melhora a performance de modelos induzidos pelo algoritmo escolhido
4. Para o conjunto de dados escolhido, o algoritmo de árvore de decisão induz um modelo com melhor performance do que o induzido pelo algoritmo de k -NN

após realizar o processo de otimização de hiperparâmetros (testar ao menos 100 conjuntos de hiperparâmetros para cada algoritmo)

Comentários adicionais: não é necessário realizar um *teste de hipótese* estatístico nesta quest. Pode testar as hipóteses apenas checando o que ocorre e elaborando uma conclusão razoável balizada no bom-senso. Mas se quiser usar teste de hipótese estatístico fique à vontade também, só não é um requerimento. O bom uso das ferramentas e estratégias discutidas em sala é esperado e será considerado na pontuação.

Pontuação máxima: 25 pontos para cada hipótese testada (máximo de 75 pontos)



Figura 1: Milu Iluminarius MLVI segurando a Grande Espada da Luz da Lua (imagem gerada por IA no site Hotpot).

2 A magia dos pipelines

ID: Modelos 5

Proponente: Haroldo Poteiro

Lore: A Guilda dos Magos é provavelmente a guilda com a maior quantidade de grandes nomes pertencentes a ela. A Guilda foi fundada pelos lendários magos e feiticeiras Albarto Dom Bolodor, Omi Nistro, Coluena Garrorvo, Glovely Terralta e aquele que ficou conhecido como O Andarilho Arco-íris. A Guilda aceita tanto novos membros de famílias com tradição nas artes arcanas (assim como o caso de Ronaldo Furão da tradicional família Furão) bem como membros vindos de famílias sem nenhuma tradição mágica (como foi o caso de Ermínia Granjeira). Mesmo tendo se isolado da sociedade, a feiticeira Ifa, a rainha do fogo, faz parte da Guilda dos Magos. Dizem por aí que na hora de escolher a sua Guilda, Ifa escolheu a Guilda dos Magos por motivos amorosos.

Objetivo: Crie pelo menos 3 pipelines diferentes e utilize eles em um conjunto de dados a sua escolha. Pelo menos um destes pipelines deve ser criado pela função `Pipeline` do `scikit-learn` (ver mais informações [aqui](#)). Pelo menos um destes pipelines deve conter 3 ou mais passos. Os pipelines devem fazer sentido.

Leitura recomendada: https://scikit-learn.org/stable/data_transforms.html

Pontuação máxima: 25 pontos.



Figura 2: Retrato artístico do Andarilho Arco-íris (imagem gerada por IA no site Hotpot).

3 Tentando algo diferente

ID: Modelos 6

Proponente: Ricardo Girado

Lore: Ricardo vem de uma família tradicional de guerreiros, mas seu coração sempre esteve em outro lugar. No fundo, ele sonhava em se tornar um bardo, em cativar o público com as suas melodias cheias de emoção e sentimentos... sempre que via uma performance da Senhorita Creizuda ou da Alfaiate Ligeirinha, sua vontade crescia ainda mais. Mesmo assim, Ricardo sabia das regras e sabia que deveria continuar sendo um guerreiro. Apesar das expectativas de seus pais e familiares, Ricardo não abriu mão da sua paixão; “nunca vou desistir de ti” repetia sempre. Quando as noites são calmas e as estrelas brilham no céu, Ricardo vai pra floresta tocar seu alaúde. Canta sobre um amor tão forte que nunca abandonaria, nunca desiludiria, nem correria ou magoaria até o fim da sua vida. As suas canções, cheias de emoção, ecoam pela floresta como uma promessa secreta à vida que ele desejava mas que nunca poderia abraçar completamente.

Objetivo: Escolha um dos seguintes algoritmos de aprendizado de máquina: Support Vector Machine, Naive Bayes ou Gaussian Processes. Sua tarefa é entender como este algoritmo funciona e reportar isso de maneira clara e didática. Induza um modelo utilizando este algoritmo nos dados que pretende usar no seu projeto de conclusão do semestre desta disciplina.

Pontuação máxima: 50 pontos.



Figura 3: Ricardo Girado (imagem gerada por IA no site Hotpot).

4 Definitivamente incerto

ID: Módulos 2

Proponente: Dedé Recicles

Lore: Dedé Recicles é um filósofo, cientista, matemático, ator, padreiro, tecelão, mestre de obras e animador de festas infantis (este último apenas nos finais de semana). As pessoas questionam como que Recicles consegue existir sendo que não parece ter tempo nem para pensar direito, considerando tanta coisa que faz. De qualquer maneira, nas horas vagas (!) Recicles busca pistas sobre o Clã dos Barrigas Rastejantes e quem seria a figura de liderança por trás deste clã... as pistas até agora apontam para algo estranho dentro das Catacumbas da Escuridão Sombria...

Objetivo: Faça um notebook didático mostrando os princípios básicos de uso do módulo `uncertainties` do Python (<https://uncertainties.readthedocs.io/en/latest/>). Este módulo é utilizado para propagação de erros.

Pontuação máxima: 15 pontos.



Figura 4: Dedé Recicles (imagem gerada por IA no site Hotpot).