

Análise de Dados em Astronomia - AGA0505 - 2025

Perguntas 1:

1. o que é um espaço de dados?
2. descreva os principais tipos e formas de dados
3. o que é uma variável aleatória?
4. o que é uma amostra?
5. o que é amostragem de uma distribuição de probabilidades?
6. qual é a diferença entre estatísticas da população e estatísticas de uma amostra?
7. o que são probabilidades de um ponto de vista frequentista e bayesiano?
8. o que é uma função de densidade de probabilidades?
9. o que é uma função de massa de probabilidades?
10. o que é uma função de distribuição cumulativa?
11. quais são os valores mínimo e máximo de uma função de distribuição de probabilidades cumulativa?
12. o que é a função quantil?
13. o que são quartis?
14. o que é a estatística descritiva? qual é seu objetivo?
15. o que é uma estatística?
16. o que é uma estatística robusta?
17. o que é viés (*bias*) estatístico?
18. como as estatísticas de uma variável se relacionam com sua distribuição de probabilidades?
19. como se define o valor esperado de uma função $f(x)$ com respeito a uma distribuição de probabilidades $P(x)$?
20. dê exemplos de estatísticas de posição de distribuições.
21. dê exemplos de estatísticas de largura de distribuições.
22. dê exemplos de estatísticas de forma de distribuições.
23. desenhe uma distribuição com assimetria (*skewness*) negativa.

Perguntas 2:

1. descreva alguns tipos de distribuições de probabilidades
2. o que são probabilidades condicionais? como são representadas?
3. o que são probabilidades conjuntas? como são representadas?
4. escreva a probabilidade conjunta de duas variáveis x e y dada uma variável z .
5. escreva a probabilidade de uma variável x condicionada a duas variáveis conjuntas y e z .
6. o que é a regra da soma das probabilidades?
7. o que é a regra do produto?
8. derive o teorema de Bayes a partir da definição de probabilidade conjunta de duas variáveis.
9. mostre que o produto de duas proposições independentes é o produto das probabilidades de cada uma.
10. como se define a marginalização de uma variável contínua?
11. como se define a marginalização de uma variável discreta?
12. o que é a lei das probabilidades totais?
13. o que é uma distribuição gaussiana bivariada?
14. como se pode determinar a função de distribuição de probabilidades de uma quantidade $y = f(x)$ conhecendo-se a função de distribuição de probabilidades $P(x)$?
15. **numa sala de aula 40% dos alunos sabem R e python e 60% sabem python; qual é a probabilidade de um estudante que conhece Ppython conhecer também R?**
16. **em um aglomerado estelar, cerca de 70% de suas estrelas têm um sistema planetário; sabe-se que 50% das estrelas com sistemas planetários são de alta metalicidade, enquanto que esta percentagem entre as que não têm sistemas planetários é de 30%. qual é a probabilidade de que uma estrela escolhida ao acaso tenha um sistema planetário e seja de baixa metalicidade?**

Perguntas 3:

1. o que são distribuições de probabilidades contínuas e discretas?

2. o que é espaço amostral?
3. quais são os parâmetros de uma distribuição gaussiana?
4. como é definida a precisão em uma distribuição gaussiana?
5. como é definida a distribuição cumulativa de uma distribuição gaussiana?
6. o que é o teorema do limite central?
7. porque muitas vezes se supõe que os erros de uma medida são gaussianos?
8. o que é a distribuição binomial? quando é usada?
9. **se 60% das estrelas são binárias, qual é a probabilidade de que uma amostra aleatória de 5 estrelas contenha 4 binárias?**
10. quanto valem a média e a variância de uma distribuição binomial?
11. para que tipo de distribuição a binomial converge para grande número de objetos?
12. o que é a distribuição de Poisson? quando é usada?
13. quanto valem a média e a variância de uma distribuição de Poisson?
14. para que tipo de distribuição a poissoniana converge para grande número de objetos?
15. o que é a distribuição exponencial? quando é usada?
16. quanto valem a média e a variância de uma distribuição exponencial?
17. se em uma galáxia uma SN explode a cada 10 anos, qual a probabilidade de que demore 100 anos para uma explodir?
18. defina a distribuição uniforme.
19. **calcule a distribuição cumulativa de uma variável uniformemente distribuída entre 0 e 1. calcule também a média e o desvio padrão dessa variável.**
20. o que é a função beta? ela se aplica a que tipo de variável?
21. dê exemplos de distribuições em lei de potência de interesse em Astronomia.
22. o que é necessário para que uma função de duas variáveis seja considerada uma função de distribuição de probabilidades bivariada?
23. o que são distribuições marginais?
24. como é definida a covariância entre duas variáveis?
25. qual é a covariância entre duas variáveis independentes?

26. como é definida a distribuição gaussiana bivariada?
27. escreva a função de distribuição da gaussiana bivariada em notação vetorial.
28. como se define o coeficiente de correlação de Pearson?

Perguntas 4:

1. o que é amostragem de uma distribuição de probabilidades?
2. para que serve o método de Monte Carlo?
3. como se pode amostrar uma distribuição de probabilidades a partir de sua distribuição cumulativa?
4. **ilustre como funciona o método de Monte Carlo amostrando $P(x) = \exp(-x)$, com $x > 0$.**
5. como a variância de um resultado obtido por Monte Carlo varia com o número de simulações?
6. se quero diminuir por um fator 2 o desvio padrão de uma variável amostrada por Monte Carlo, de quanto devo aumentar o tamanho da amostra?
7. no exemplo do universo homogêneo, porque o raio é amostrado como $R\gamma^{1/3}$, onde γ é um número aleatório uniformemente distribuído entre 0 e 1?
8. que tipo de distribuição radial de densidade se obteria se amostrassemos o raio como $R\gamma$?
9. no exemplo do universo homogêneo, como pode se amostrar θ ?
10. no exemplo do universo homogêneo, como pode se amostrar ϕ ?
11. no exemplo do universo homogêneo, como o número cumulativo de galáxias varia com r ?
12. para que serve MCMC?
13. o que é um processo de Markov?
14. o que é a função proposta em MCMC?
15. descreva o algoritmo de Metropolis-Hastings.
16. o que é o “burn-in” numa simulação MCMC?
17. como se pode avaliar se uma simulação MCMC convergiu?

Perguntas 5:

1. porque o teorema de Bayes é útil em análise de dados?
2. descreva o teorema de Bayes, apontando o significado do prior, da verossimilhança, da evidência e do posterior.
3. o que é a evidência? em que tipo de problema ela é útil?
4. para que serve a distribuição preditiva de um novo dado?
5. o que são os intervalos de credibilidade de uma variável?
6. como o teorema de Bayes pode ser usado para inferência dos parâmetros de um modelo?
7. como se estima incertezas em parâmetros a partir do posterior?
8. como o teorema de Bayes pode ser usado para comparar dois modelos?
9. o que é estimativa de ponto de um parâmetro e como ela se compara com uma estimativa bayesiana?
10. comente como a escolha de um prior informativo (gaussiano) afeta a estimativa da média de uma gaussiana.
11. o que é o método da máxima verossimilhança?
12. quando o máximo do posterior é igual ao máximo da verossimilhança?
13. quando o máximo do posterior é diferente do máximo da verossimilhança?
14. compare os métodos frequentistas e bayesianos de estimativa de parâmetros.
15. o que é um *nuisance parameter*?
16. o que são hiperparâmetros?
17. dê exemplos onde estimativas bayesianas e frequentistas divergem.
18. como funciona um modelo gerativo?
19. em que casos modelos gerativos são úteis?
20. qual é o objetivo do algoritmo ABC?
21. descreva o algoritmo ABC.

Perguntas 6:

1. como se comparam dois modelos com o teorema de Bayes?

2. o que são *odds*?
3. o que é a escala de Jeffreys?
4. o que é a aproximação de Laplace?
5. o que é o fator de Bayes?
6. o que é o fator de Occam?
7. como se define o BIC e o AIC? para que servem?
8. dê um exemplo do uso do BIC.

Perguntas 7:

1. para que serve o algoritmo do gradiente descendente?
2. como funciona o algoritmo do gradiente descendente?
3. suponha que tenhamos n dados independentes $D = \{x_i, y_i, \sigma_i\}$, com erros gaussianos σ_i , que queremos ajustar com uma função com m parâmetros w : $y = f(x; w)$; escreva a verossimilhança do modelo.
4. o que é o ' χ^2 reduzido'?
5. que valor se espera, mais ou menos, para o χ^2 reduzido de um bom ajuste?
6. como os erros são estimados no método da máxima verossimilhança?
7. o que são simulações de *bootstrap*? para o que são usadas?
8. como funciona o algoritmo de *bootstrap*?
9. como se pode determinar intervalos de confiança por *bootstrap*?
10. o que significa um intervalo de confiança de 90% em torno da média de uma variável aleatória?
11. o que é a regressão linear ordinária (OLS)
12. o que é um modelo linear?
13. como um modelo do tipo $y = a10^{bx}$ pode ser transformado num modelo linear?
14. como um modelo do tipo $y = ax^b$ pode ser transformado num modelo linear?
15. porque o modelo $y = a + bg(x)$, com parâmetros a e b e $g(x)$ uma função qualquer de x , pode ser descrito por um modelo linear?
16. o que é o valor p ? como ele é usado em testes de hipóteses?

17. o que é uma hipótese nula?
18. considere a distribuição $P(x) = 2x$ ($0 < x < 1$). Qual é a probabilidade de se ter $x > 0.9$?
19. como se testa se duas distribuições têm a mesma média?
20. para que serve o teste de Kolmogorov-Smirnov?
21. o que o coeficiente de correlação entre duas variáveis mede?
22. como é definido o coeficiente de correlação de Pearson?
23. o que é o coeficiente de correlação de Spearman? que vantagem ele tem em relação ao coeficiente de correlação de Pearson?

Perguntas 8:

1. como se relacionam o Deep Learning, o Machine Learning e a Inteligência Artificial?
2. o que é o espaço de dados? dê um exemplo.
3. o que são *features*?
4. considere o problema de classificação de objetos em uma imagem: comente sobre como abordar o problema com programação clássica e com aprendizado de máquina.
5. o que é aprendizagem em ML?
6. que tipo de procedimento numérico é equivalente a aprendizagem?
7. comente a relação entre algoritmos de ML e funções.
8. o que é aprendizado supervisionado?
9. o que é regressão?
10. o que é classificação?
11. o que é aprendizado não-supervisionado?
12. o que é aprendizado com reforço?
13. o que são modelos paramétricos?
14. o que são modelos não-paramétricos?
15. o que é a generalização em ML?
16. o que é *under-fitting*?

17. o que é *over-fitting*?
18. qual é o objetivo dos métodos de estimativa de densidades?
19. em que condições uma distribuição pode ser considerada uma distribuição de probabilidades?
20. no caso de histogramas, porque é difícil se saber o tamanho ótimo de um bin?
21. o que é e para que se usa o KDE?
22. o KDE é considerado um método paramétrico ou não-paramétrico?
23. quais são os parâmetros do KDE?
24. como a escolha da largura de banda do KDE afeta estimativas?
25. como se pode escolher a largura de banda?
26. como você avalia o efeito do tipo de kernel nas estimativas?
27. qual é o objetivo da análise de agrupamento (*cluster analysis*)?
28. a análise de agrupamento é uma técnica supervisionada ou não? porque?
29. porque se definem distâncias no espaço de dados?
30. o que é o algoritmo k-means? qual é seu principal parâmetro?
31. o que é PCA? como funciona?
32. para quê o PCA é útil?
33. o que são os autovetores e autovalores da matriz de correlação?
34. o que é o *embedding*?
35. o que é *locally linear embedding*?
36. como se pode avaliar a qualidade da compressão de informação com a biblioteca dimRed?

Perguntas 9:

1. como se relacionam o Deep Learning, o Machine Learning e a Inteligência Artificial?
2. o que é o espaço de dados? dê um exemplo.
3. o que são *features*?

4. considere o problema de classificação de objetos em uma imagem: comente sobre como abordar o problema com programação clássica e com aprendizado de máquina.
5. o que é aprendizagem em ML?
6. comente a relação entre algoritmos de ML e funções.
7. o que é aprendizado supervisionado?
8. o que é regressão?
9. o que é classificação?
10. o que é aprendizado não-supervisionado?
11. o que é aprendizado com reforço?
12. o que são modelos paramétricos?
13. o que são modelos não-paramétricos?
14. o que é a generalização em ML?
15. o que é *under-fitting*?
16. o que é *over-fitting*?
17. qual é o objetivo dos métodos de estimativa de densidades?
18. dê exemplos de métodos de estimativa de densidades.
19. o que é e para que se usa o KDE?
20. o KDE é considerado um método paramétrico ou não-paramétrico?
21. como a escolha da largura de banda do KDE afeta estimativas?
22. como se pode escolher a largura de banda?
23. qual é o objetivo da análise de agrupamento (*cluster analysis*)?
24. a análise de agrupamento é uma técnica supervisionada ou não? porque?
25. o que é o algoritmo k-means? qual é seu principal parâmetro?
26. o que é PCA? como funciona?
27. para quê o PCA é útil?
28. o que é o *embedding*?
29. quando métodos de redução de dimensionalidade podem ser úteis?

Perguntas 10:

1. o que são redes de neurônios artificiais (RNAs)?
2. quais são as vantagens das RNAs?
3. dê exemplos de diferentes tipos de arquitetura de RNAs.
4. que tipos de função de ativação são normalmente usados?
5. como é definida a função sigmoide?
6. qual é a arquitetura típica de um perceptron multi-camadas (MLP)?
7. porque os MLPs são também denominados redes densamente conectadas?
8. como é feita a inferência com uma RNA?
9. o que é a regressão logística?
10. o que é *back-propagation*?
11. o que diz o teorema da universalidade?
12. o que são redes convolucionais?
13. com que tipo de dados as redes convolucionais são mais usadas?
14. o que são redes pré-treinadas?
15. o que é *transfer learning*?
16. como se controla o overfitting em redes convolucionais?

Perguntas 11:

1. o que é uma série temporal?
2. quais são os 3 tipos de fenômenos que são dependentes do tempo?
3. o que são processos periódicos ou determinísticos?
4. o que são processos transientes ou explosivos?
5. o que são processos semi-periódicos, estocásticos ou caóticos?
6. comente algumas diferenças entre a análise de séries temporais em astronomia e em outras áreas (como economia).
7. o que é a função de autocorrelação de uma série temporal?

8. porque às vezes se faz filtragem de uma série temporal?
9. discuta alguns problemas com a filtragem.
10. qual é o objetivo da análise de Fourier?
11. o que é o espectro de potências (periodograma)?
12. discuta alguns problemas com a análise harmônica
13. o que são wavelets?
14. o que são modelos autoregressivos?
15. discuta algumas aplicações de aprendizado de máquina na análise de uma série temporal.