

Lista de Exercícios 04

Professores: Erickson, Fabricio e Renato

Política da Disciplina: Leia todas as instruções abaixo cuidadosamente antes de começar a resolver a lista, e antes de fazer a submissão.

- As questões podem ser discutidas entre até três alunos (conjuntos disjuntos). Os nomes dos colegas precisam ser incluídos na submissão.
- A submissão deve ser feita em formato PDF através do Moodle, mesmo que tenham sido resolvidas a mão e escaneadas.
- Todas as soluções devem ser justificadas.
- Todas as fontes de material precisam ser citadas. O código de conduta da UFMG será seguido à risca.

Problema 1: Considere o conjunto de dados abaixo com as notas de 5 estudantes em 4 disciplinas. Calcule a matriz de covariância. (Dica: Para saber se devemos usar $\frac{X^T X}{n-1}$ ou $\frac{X X^T}{n-1}$, lembre-se de que a matriz final deve ter ordem igual ao número de atributos.)

Estudante	GAAL	PDS1	Cálculo 1	ALC
1	90	80	60	95
2	65	75	90	70
3	40	90	60	55
4	80	60	59	75
5	60	100	80	80

Problema 2: Seja X um conjunto de dados já centralizado. Sabendo que $C_x = X^T X / (n - 1)$ é a matriz de covariância de X , mostre, algebricamente, como o PCA de X pode ser obtido a partir de seu SVD.

Problema 3: Considere os pontos a seguir:

x	2.0	3.5	4.0	5.1	7.0
y	2.2	2.0	3.0	6.0	5.0

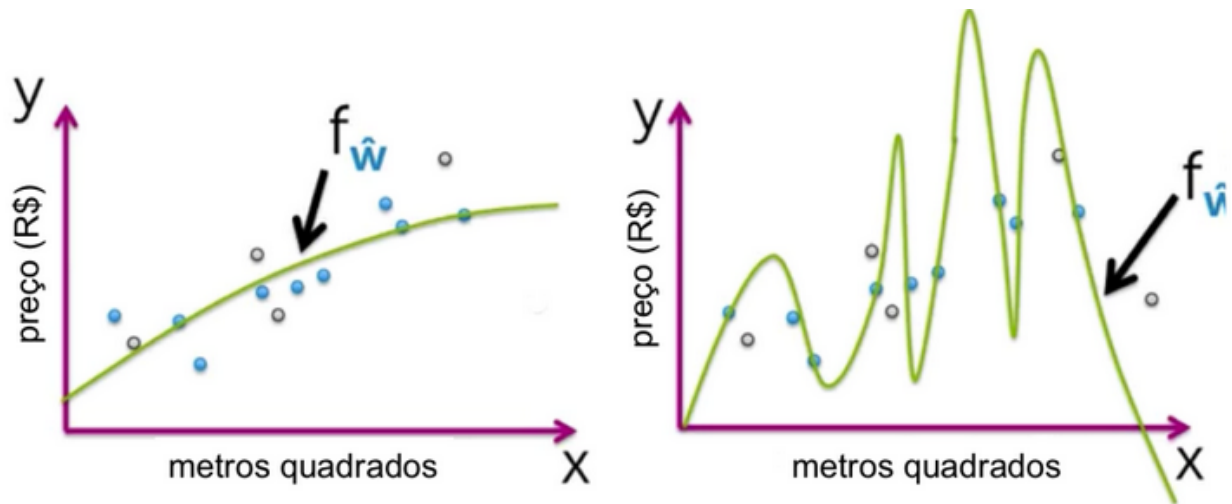
Usando o método dos quadrados mínimos, encontre os parâmetros da regressão linear simples $f(x) = \beta_0 + \beta_1 x$. **Atenção:** você não pode resolver esta questão usando uma função de biblioteca que retorne os coeficientes da regressão diretamente a partir de x e y .

Problema 5: Deseja-se usar a regressão polinomial $f(x_i) = w_0 + w_1 x_i + w_2 x_i^2 + \dots + w_p x_i^p$ para estimar a relação entre a metragem (em m^2) de um imóvel e o seu preço em um bairro de Belo Horizonte. As figuras abaixo ilustram (não são uma representação exata) os resultados obtidos para $p = 3$ e $p = 8$, respectivamente, a partir dos pontos em azul.

Responda:

- Qual das regressões possui o menor desvio considerando-se apenas os pontos em azul?
- Qual das regressões possui o menor desvio considerando-se os pontos azuis e os pontos brancos?

- Com base nesta última resposta, qual dos valores de p é mais adequado e por quê?



Problema 6: Assinale V para verdadeiro ou F para falso e justifique:

- () Dado um conjunto de dados centralizado X , para obter uma representação de X em k dimensões via PCA, podemos utilizar o SVD truncado de X de posto k .
- () A direção da PC_1 de um conjunto de dados bidimensional, em que cada ponto i tem coordenadas (x_i, y_i) , é a mesma da reta $f(x)$ obtida pela regressão linear dos dados.