

## **Lista 2**

**José Elias cangombe**

**Matrícula - 799**

**1. Qual técnica de regressão linear você usaria se tivesse um conjunto de treinamento com milhares de features? Explique por quais razões você utilizaria esta técnica.**

R: Usaria a técnica de regressão linear gradiente descendente estocástico, porque apresenta iteratividades que permite obter uma convergência mais rápida e diferente do algoritmo em batelada, onde este modelo utiliza apenas uma parte deste conjunto de treinamento de forma aleatória.

**2- Suponha que as features (i.e., atributos) do seu conjunto de treinamento tenham escalas muito diferentes. Qual técnica de regressão linear pode sofrer com isso e como? O que pode ser feito para mitigar este problema?**

R: É a técnica de escalonamento de feature e em algumas situações, alguns atributos acabam sendo dominantes sobre os demais no sentido de que exercem grande influência sobre o erro cometido pelo modelo. E para mitigar os problemas será necessário escalonar a variação de todos os atributos do erro quadrático médio.

**3- Suponha que você use o gradiente descendente em batelada e plote o erro de cada época. Se você perceber que o erro aumenta constantemente, o que provavelmente está acontecendo? Como você pode consertar isso?**

R: Sempre que o erro aumenta constantemente, é sinal de que o passo de aprendizagem implementado esteja demasiadamente grande de modo que o algoritmo diverge. E para consertar este problema será necessário um ajuste no passo de aprendizagem através de uma análise gráfica da função de custo e do número de iterações.

**4. Entre os algoritmos baseados no gradiente descendente (GD) que discutimos (batch, Estocástico e mini-batch), qual deles chega mais rapidamente à vizinhança da solução ótima? Qual deles realmente converge? O que você pode fazer para que os outros também converjam?**

R: O que chega mais rápido é o gradiente descendente estocástico.

O gradiente descendente em batelada (batch) caminha diretamente para o mínimo. E ele é o que realmente converge.

E para que os outros algoritmos possam convergir é necessário modificar o passo de aprendizagem.