

TP555 - AI/ML

Nome: Mayomona Lando Filipe

Matricula: 836

Lista de Exercícios #2

Regressão Linear

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: Para um treinamento com vários features, uma das técnicas mais adequadas seria a técnica de regressão linear com gradiente descendente estocástico.

Utilizaria o gradiente descendente estocástico devido ao fato de não usar todos os dados do conjunto de treinamento, mas sim usar uma fração dos dados. Esse subconjunto dos dados é obtido com amostragem dos dados originais.

2. Suponha que as features (Classificações) (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: A técnica a usar é o escalonamento de features. O problema ocorre quando as features têm escalas diferentes, devido à grande diferença de magnitude entre os atributos, isso gera problemas com todo algoritmo que se baseia no cálculo de distância na fase do aprendizado.

A forma de mitigar este problema é fazer com que todos os atributos tenham a mesma escala: escalonamento min-max (também conhecido como normalização) e a padronização, ou seja a variação de todos os atributos deve ser escalonada para que cada atributo contribua com a mesma importância/peso para o cálculo da distância (ou seja, do erro quadrático médio), de forma que eles tenham a mesma faixa de valores.

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: Quando o erro aumenta constantemente, isso pode ser um indicativo de que o passo de aprendizagem utilizado esteja excessivamente grande, fazendo o algoritmo divergir. Para que o problema seja solucionado é necessário ajustar o passo de aprendizagem adequadamente através de uma análise do gráfico da

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: Tanto no gradiente descendente do quanto no descendente estocástico você atualiza um conjunto de parâmetros de maneira iterativa para minimizar uma função de erro.

Enquanto estiver no gradiente descendente, você precisará executar todas as amostras em seu conjunto de treinamento para fazer uma única atualização para um parâmetro em uma iteração específica, no gradiente descendente estocásticos, por outro lado, você usa apenas uma ou subconjunto de amostra de treinamento do seu conjunto de treinamento para fazer a atualização para um parâmetro em uma iteração específica. Se nos usar subconjunto, será chamado de gradiente descendente estocástico Mini-batch.

Portanto, se o número de amostras de treinamento for grande, na verdade muito grande, o uso do gradiente descendente poderá demorar muito, pois em todas as iterações, quando você está atualizando os valores dos parâmetros, está executando o conjunto de treinamento completo. Por fim, o uso do gradiente descendente estocástico será mais rápido chegar à vizinhança da solução ótima, porque ele usa apenas uma amostra de treinamento e ela começa a melhorar a partir da primeira amostra.

O gradiente descendente estocástico costuma convergir muito mais rapidamente em comparação ao gradiente descendente, mas a função de erro não é tão minimizada quanto no caso do gradiente descendente. Geralmente, na maioria dos casos, a aproximação aproximada que ele obtém no gradiente descendente estocástico para os valores dos parâmetros é suficiente porque eles atingem os valores ótimos e continuam oscilando lá.

