A blue and black logo

Description automatically generatedAprendizagem 2024/2025

**Homework III – Group 034**

(ist199417, ist1106481)

**I. Pen-and-paper**

**1)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 1 | 1 | 1 | 1.25 |
|  | 1 | 3 | 3 | 7 |
|  | 3 | 2 | 6 | 2.7 |
|  | 3 | 3 | 9 | 3.2 |
|  | 2 | 4 | 8 | 5.5 |

Definimos a matriz X e vetor z:

Usando numpy para os cálculos, temos:

**2)**

Utilizando a mesma matriz e o mesmo vetor do exercício 1 e numpy para os cálculos, temos:

O fator de regularização ajuda a combater o overfitting, ao diminuir a magnitude de coeficientes sobrestimados ou subestimados. É precisamente isso que vemos a acontecer. Os coeficientes da pergunta anterior são (3.31593, 0.11372) enquanto que os coeficientes com regularização são (1.81809, 0.32376). O coeficiente 3.31593 foi diminuído e o coeficiente 0.11372 foi aumentado.

**3)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | 2 | 2 | 4 | 0.7 |
|  | 1 | 2 | 2 | 1.1 |
|  | 5 | 1 | 5 | 2.2 |

|  |  |  |
| --- | --- | --- |
|  | OLS | Ridge |
|  | 3.42965 | 2.14184 |
|  | 3.65708 | 2.78936 |
|  | 3.99823 | 3.76064 |
|  | 4.33938 | 4.73191 |
|  | 4.22566 | 4.40816 |
|  | 3.77080 | 3.11312 |
|  | 3.54336 | 2.46560 |
|  | 3.88451 | 3.43688 |

Treino

Teste

Com os dados do enunciado, obtemos e, com os dados desta tabela, obtemos

(Treino, OLS)

(Treino, Ridge)

(Teste, OLS)

(Treino, Ridge)

O resultado obtido foi de encontro com o esperado. Ao adicionarmos o fator de regularização, estamos a tentar diminuir o overfitting, podendo o modelo ser pior no conjunto de treino, para depois vir a ser melhor nos dados de teste. Como vemos, no grupo de treino, o RMSE sem regularização foi de 2.02650, enquanto que o RMSE com regularização foi de 2.15354. Portanto, de facto, no grupo de treino, a regularização levou a um aumento do erro da previsão em relação ao valor verdadeiro de ynum. Agora no grupo de teste, o RMSE sem regularização foi de 2.46559, enquanto que o RMSE com regularização foi de 1.75281. Ou seja, o modelo com regularização teve um resultado muito mais satisfatório no grupo de teste do que o modelo sem regularização. Foi até o menor erro entre todos os erros calculados!

**4)**

(Usámos numpy para fazer os cálculos)

Pesos e Bias atualizados:

Forward pass:

Backpropagation:

Resposta:

**II. Programming and critical analysis**

A graph showing different models

Description automatically generated**5)**

**6)**

Numa regressão linear, o modelo assume uma relação linear entre as features de entrada e a variável target. Um MLP sem ativações comporta-se essencialmente como um modelo linear, já que cada layer passa a weighted sum sem realizar qualquer transformação. Observando o gráfico, vemos que corrobora a nossa tese, o boxplot MAE da 'Linear Regression' é praticamente igual ao do MLP sem ativações.

Sem funções de ativação, as redes neuronais estariam limitadas a transformações lineares, tornando-as incapazes de aprender e modelar dados complexos. Ao introduzir não-linearidade, as funções de ativação permitem que as redes neuronais aproximem funções arbitrárias, tornando-as ferramentas poderosas para tarefas como o reconhecimento de imagens, o processamento de linguagem natural e muitas outras.

Como podemos ver pelo boxplot, ao usar a função de ativação ReLu, o modelo apresentou um desempenho signficamente melhor, tendo-se ajustado melhor a estes dados.

**7)**

**A graph of a test

Description automatically generated with medium confidence**

Melhores parametros**:**

-alpha: 0.01

- batch\_size: 32

- learning\_rate\_init: 0.001

Ao observarmos este mapa de calor, podemos começar por ver que à medida que o learning rate aumenta, o MAE também aumenta. Podemos ver isto fixando as outras duas variáveis (alpha e batch size), concluindo que o melhor valor de learning rate será o mínimo (0.001).

Os valores de alpha podem produzir overfitting (valor baixo) ou underfitting (valor alto), dependendo da combinação.

Ao aumentar o batch size, geralmente é mais estável o treino, levando a valores de MAE mais baixos. Observando o gráfico, vemos que não parece haver problema quando a learning rate é baixa, começando-se a notar para learning rates de 0.01 e 0.1.