

Exercício de Fixação de Conceitos

EFC2 - Questão 2

Nome: Guilherme Rosa

RA: 157955

1. Separação das amostras de treinamento nos conjuntos de treinamento e validação

As 60.000 amostras de treinamento foram permutadas pseudo-aleatoriamente e divididas em dois conjuntos para utilizar a técnica de validação *holdout*, cuja finalidade é aumentar a capacidade de generalização da ELM. A divisão das amostras foi feita da seguinte forma:

- 80% das amostras foram agrupadas no conjunto de treinamento (48.000)
- 20% das amostras foram agrupadas no conjunto de validação (12.000)

Em seguida foi verificado a proporção de amostras por classe em ambos os conjuntos, de modo a verificar o balanceamento e representatividade das classes. As porcentagens de cada classe variaram entre 8.99% a 11.29% no conjunto de treinamento e 9.2% a 11.02% no conjunto de validação. Nota-se que todas as classes estão representadas de forma equilibrada, o que leva a classificadores menos enviesados.

2. Máquina de Aprendizado Extremo (ELM)

Conforme especificado no enunciado da questão, foi implementado uma máquina de aprendizado extremo com 500 neurônios na camada intermediária cujo os pesos foram definidos aleatoriamente a partir de uma função de distribuição normal com média nula e desvio padrão de 0.2. Além disso, foi definida função de ativação do tipo tangente hiperbólica para cada um dos neurônios.

A implementação feita da ELM também permite o usuário modificar os parâmetros da distribuição normal, utilizar uma distribuição uniforme para definição aleatória dos pesos e ainda utilizar função de ativação ReLU no lugar da tangente hiperbólica. O treinamento dessa nova máquina é feito com o algoritmo desenvolvido para o classificador linear da questão 1.

O código pode ser visitado em https://github.com/guilherme-rosa/IA353_RedesNeurais no arquivo *EFC1 - Parte 2.ipynb* do diretório *EFC1*.

3. Busca inicial pelos melhores coeficientes de regularização

A busca inicial pelo melhor coeficiente de regularização λ foi feita utilizando dois critérios de desempenho: erro quadrático médio e a taxa de erro de classificação. Essa busca foi feita considerando os seguintes valores para λ :

$$\lambda = \{2^{-20}, 2^{-8}, \dots, 2^{+8}, 2^{+10}, \dots, 2^{18}, 2^{20}\}$$

As figuras 1 e 2 apresentam os gráficos semilog das métricas de desempenho em função do coeficiente de regularização. Os melhores coeficientes da busca inicial estão apresentados na primeira linha da Tabela 1.

Tabela 1: Valores dos melhores coeficientes de regularização, antes e após o refinamento, considerando o erro quadrático médio e taxa de erro de classificação

	Melhor λ para o erro quadrático médio	Melhor λ para a taxa de erro de classificação
Busca inicial	128	64
Busca refinada	141,6364	68,0404



Figura 1: Gráfico semilog do erro quadrático médio em função do λ na busca inicial.

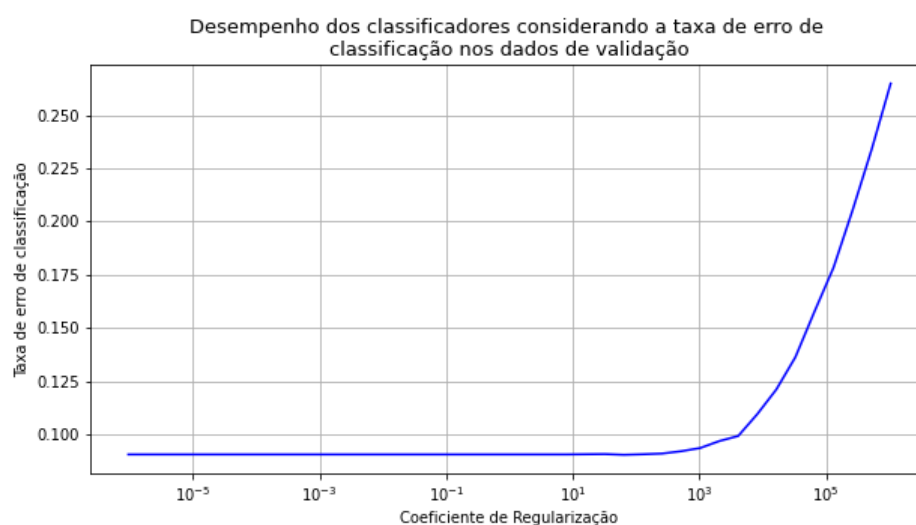


Figura 2: Gráfico semilog da taxa de erro de classificação em função do λ na busca inicial.

4. Busca refinada pelo melhor λ para o erro quadrático médio:

A busca refinada pelo coeficiente λ foi feita em duas etapas. Primeiro, uma busca foi feita considerando 100 valores linearmente espaçados no intervalo $[78, 178]$, o qual resultou num valor de 141.63636 para o coeficiente. Em seguida, foi feita um novo refinamento na busca, agora considerando 100 valores linearmente espaçados no intervalo $[163.63, 146.63]$.

As Figura 3 e 4 apresentam os gráficos semilog do erro quadrático médio em função do coeficiente para as duas etapas de refinamento. O valor do melhor coeficiente obtido está apresentado na segunda linha da Tabela 1.

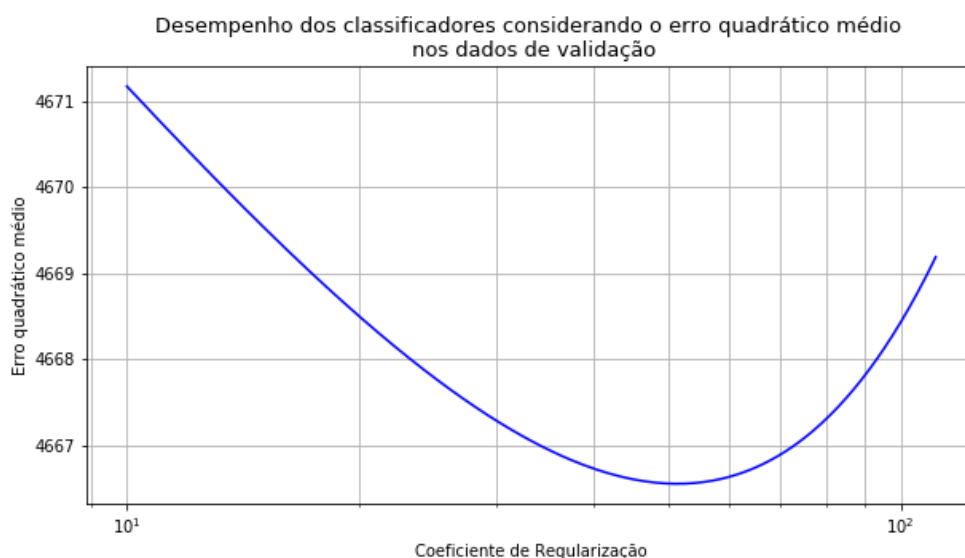


Figura 3: Gráfico semilog do erro quadrático médio em função do λ na primeira parte da busca refinada.

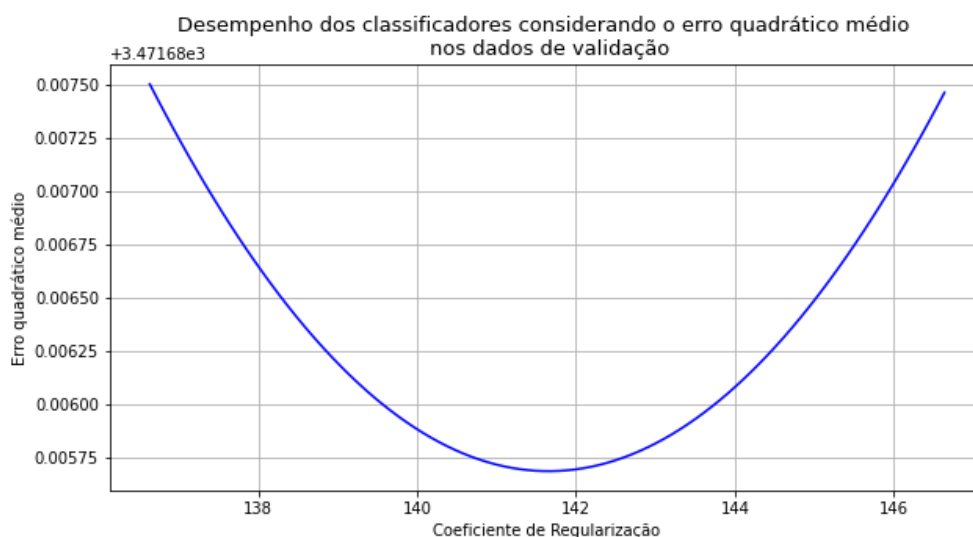


Figura 4: Gráfico semilog do erro quadrático médio em função do λ na segunda parte da busca refinada.

5. Busca refinada pelo melhor λ para a taxa de erro de classificação:

Assim como na Seção 4 a busca refinada pelo coeficiente λ foi feita em duas etapas. Na primeira foi investigado 100 valores linearmente espaçados no intervalo $[14, 114]$ e em seguida foi investigado mais 100 valores linearmente espaçados, agora no intervalo $[60, 80]$. As Figuras 5 e 6 apresentam os gráficos semilog do erro quadrático médio em função do coeficiente. O valor do melhor coeficiente obtido está apresentado na segunda linha da Tabela 1.



Figura 5: Gráfico semilog da taxa de erro em função do λ na primeira etapa da busca refinada.

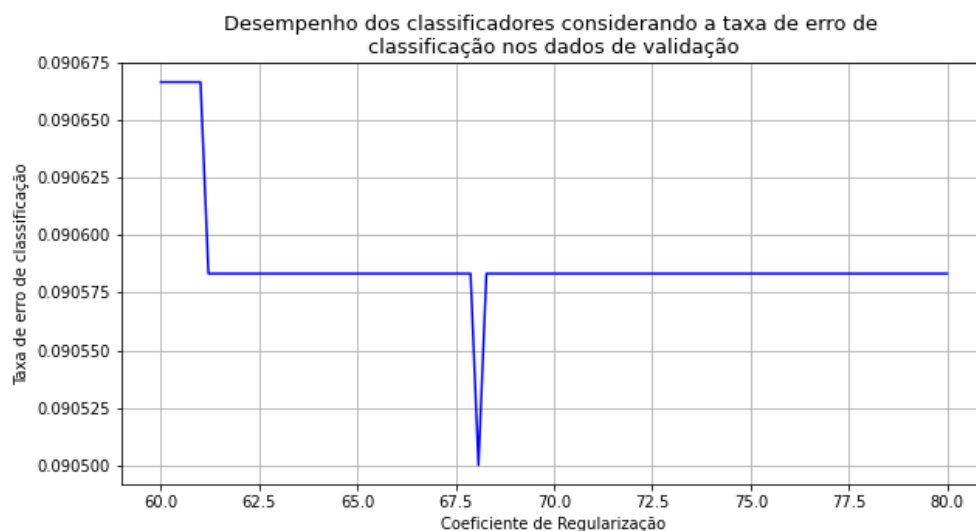


Figura 6: Gráfico semilog da taxa de erro em função do λ na segunda etapa da busca refinada.

6. Modelo Final da Máquina de Aprendizado Extremo

- Respostas referentes a questão 2.1:

Para a implementação do modelo final da ELM utilizamos o coeficiente de regularização da busca refinada associado a menor taxa de erro de classificação. Em seguida, treinamos novamente o modelo com todas as 60.000 amostras de treinamento e aplicamos o modelo resultante nos dados do próprio conjunto de treinamento.

Tabela 2: Métricas de desempenho do classificador linear final com $\lambda = 965,8832$

Parâmetro/Métrica	Valor
Erro quadrático médio	17013,9933
Taxa de erro de classificação	8,847%
Taxa de acerto (acurácia global)	91,153%

A Tabela 2 apresenta algumas informações do modelo final da ELM. Nota-se que a taxa de acerto ou acurácia global do classificador é de 91,153%, sendo este valor calculado pela razão entre o número de amostras classificadas corretamente e o número total de amostras. A Figura 7 apresenta a matriz de confusão do modelo e a Figura 8 apresenta alguns exemplos de imagens do conjunto de treinamento classificados incorretamente.

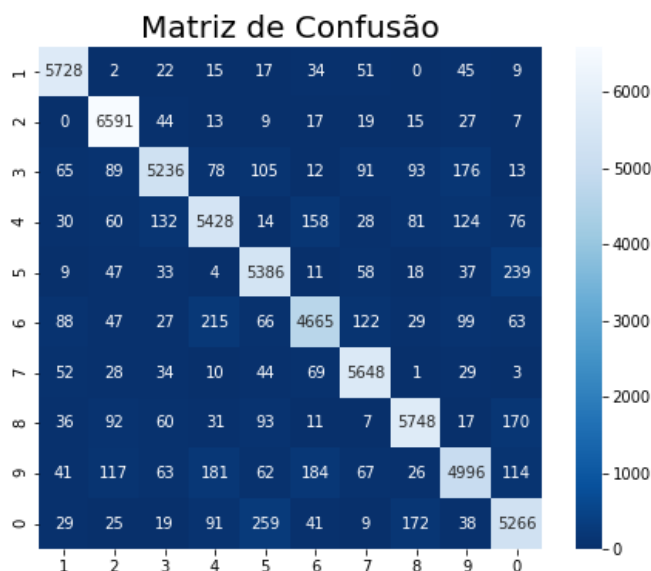


Figura 7: Matriz de confusão da ELM junto aos dados de treinamento.

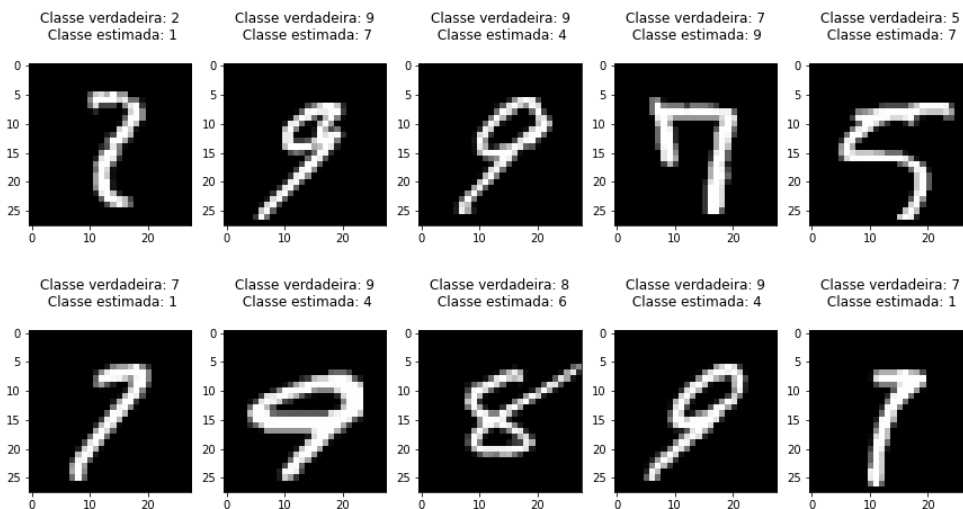


Figura 8: Exemplos de imagens classificadas incorretamente.

- Resposta referente a questão 2.2:

O ganho de desempenho da Máquina de Aprendizado Extremo em comparação ao Classificador Linear deve-se a presença da camada intermediária de neurônios que realiza transformações não-lineares nos atributos das amostras de entrada, ou seja, cada neurônio da camada intermediária é responsável por criar um mapeamento não-linear que será posteriormente combinado pelos neurônios da camada de saída para estimar a classe das amostras. Outro fator que explica o ganho de desempenho da ELM é o aumento da flexibilidade do modelo.

A diminuição no tempo de execução da ELM em comparação ao Classificador Linear deve-se a redução nas dimensões das matrizes envolvidas na resolução do problema de quadrados mínimos regularizado. No caso do Classificador Linear, o problema envolve uma matriz de parâmetros ajustáveis de dimensão 785×10 e uma matriz de amostras de entrada de dimensão $N \times 785$ (incluindo o termo de bias). No caso da ELM, o problema de otimização envolve uma matriz de amostras "transformadas" de dimensão $N \times 501$ (incluindo o termo de bias) e uma matriz de parâmetros ajustáveis de dimensão 501×10 . Como a solução de quadrados mínimos regularizado envolve inversão de matrizes, a redução das dimensões leva a um menor esforço computacional e, consequentemente, a um menor tempo de execução.

- Resposta referente a questão 2.3:

A Tabela 3 apresenta os melhores coeficientes de regularização para o Classificador Linear e para a Máquina de Aprendizado Extremo. Pode-se observar que os valores são bem distintos, principalmente no caso dos coeficientes para a taxa de erro de classificação.

Tabela 3: Melhores coeficientes de regularização, para o Classificador Linear e a ELM, após o refinamento considerando o erro quadrático médio e taxa de erro de classificação.

	Melhor λ para o erro quadrático médio	Melhor λ para a taxa de erro de classificação
Classificador Linear	51,6181	965,8832
ELM	141,6364	68,0404

- Resposta referente a questão 2.4:

Se os pesos dos neurônios da camada intermediária forem inicializados com valores distintos a cada execução, é esperado que o coeficiente de regularização associado aos menores erros (erro quadrático médio e de classificação) sejam diferentes para cada inicialização, pois inicializações diferentes resultam em problemas diferentes.

Para verificar essa hipótese, uma rede com a mesma arquitetura, mas com outra semente de gerador pseudoaleatório, foi treinada e uma busca inicial foi realizada para determinar o coeficiente de regularização relacionado a menor taxa de erro de classificação. O valor encontrado para o coeficiente foi de 16, conforme mostrado da Tabela 5 com os dois primeiros modelos alternativos.

- Resposta referente a questão 2.5:

A Máquina de Aprendizado Extremo possui os seguintes hiperparâmetros: número de neurônios da camada intermediária, o tipo de função de ativação e o tipo de inicialização dos pesos dos neurônios (função de distribuição de probabilidade e a semente do gerador pseudoaleatório).

Para encontrar uma ELM que apresentasse um desempenho superior ao da máquina obtida seguindo o roteiro, foram feitas as modificações nos hiperparâmetros como mostra a Tabela 4.

Tabela 4: Modelos alternativos da Máquina de Aprendizado Extremo.

Modelo Alternativo	Número de Neurônios	Semente	Função de distribuição	Função de ativação
1	500	10	Normal	tanh
2	500	42	Uniforme	tanh
3	500	42	Normal	ReLU
4	1000	42	Normal	tanh
5	1000	42	Normal	ReLU

Como podemos ver na Tabela 5, foram obtidas 3 máquinas de aprendizado extremo com desempenho superior àquele obtido ao seguir o enunciado, sendo que a melhor delas apresentou uma taxa de acerto de classificação de 94,535%. Podemos concluir que:

- Aumentar o número de neurônios da camada intermediária leva ao aumento do desempenho de uma ELM.

- Tanto na ELM com 500 neurônios quanto na ELM com 1000 neurônios o desempenho foi superior quando utilizada a ReLU como função de ativação.

Tabela 5: Resultados obtidos após a busca inicial do coeficiente de regularização para os modelos alternativos da Máquina de Aprendizado Extremo.

Modelo Alternativo	Coeficiente de regularização	RMSE	Taxa de erro de classificação(%)	Taxa de acerto de classificação(%)
1	16	17458.03384	8,872	91,128
2	8	17652. 49977	9,190	90,810
3	2	16385. 75946	8,172	91,828
4	64	13567. 25774	6,030	93,970
5	256	12810.81162	5,546	94,535