



INSTITUTO SUPERIOR TÉCNICO

APRENDIZAGEM AUTOMÁTICA

3^o TRABALHO LABORATORIAL
REDES NEURONAIS

Autores :

Daniel Rosa - 90041

Guilherme Viegas - 90090

Grupo 5

1 Questões

1.1

Verifica-se que no modelo sem processo de Early Stopping o ajuste do modelo aos dados de treino é demasiado grande, observável na Figura 2 pela elevada amplitude apresentada na *loss function* ao fim de poucas épocas, ou seja, apesar de, no início a função de perda diminua também no conjunto de validação, com o aumento do número de épocas (visto não existir Early Stopping), o modelo aproxima-se demasiado aos dados de treino, o que leva a uma maior dificuldade de generalização a novos dados de teste que se introduzam, ou seja verifica-se um *overfitting* em relação aos dados de treino. Com Early Stopping o modelo, como para quando a função de perda não apresenta uma melhoria significativa ao fim de algumas iterações, isto permite a diminuição do *overfitting* dos dados de treino, observável na Figura 1 pelo reduzido valor apresentado na *loss function*. Isto pode ser comprovado também comparando os valores do *Accuracy Score* onde se obtém para o modelo com Early Stopping um valor de 0.8742 e para o modelo sem Early Stopping um valor menor, de 0.8577.

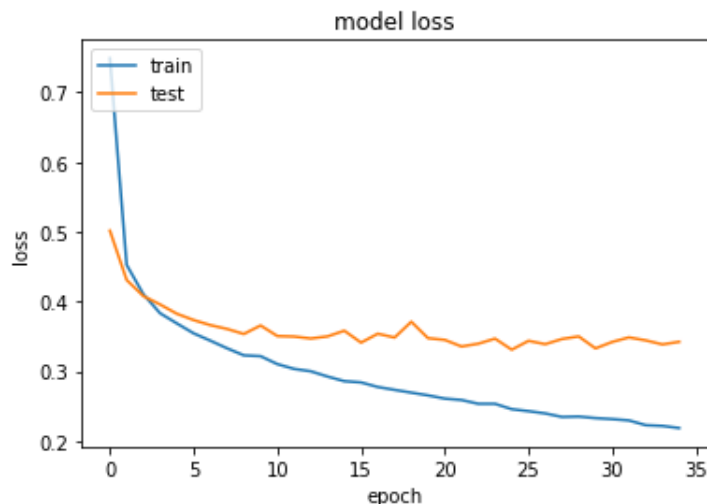


Figura 1: MLP com Early Stopping

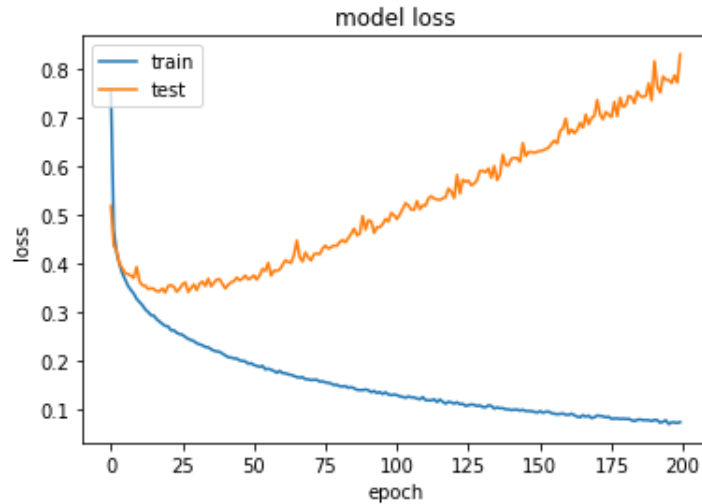


Figura 2: MLP sem Early Stopping

1.2

As MLP's usam 1 perceptrão para cada input, neste caso 1 perceptrão por píxel, e por isso a quantidade de pesos a calcular torna-se rapidamente impossível de gerir para imagens relativamente grandes. Verifica-se um número de parâmetros bastante grande, pois cada perceptrão está totalmente conectado com todos os outros da camada seguinte e anterior, o que resulta em alguma ineficiência e redundância. Em segundo lugar, as MLP's não reagem bem a translações, isto é, para uma mesma feature, como por exemplo uma bota, se aparecer no canto esquerdo de uma imagem e no direito de outra, a própria MLP vai tentar corrigir-se assumindo que a bota estará sempre numa secção da imagem. Por este motivo, as MLP's não são a melhor solução para processamento de imagens.

As CNN's representam internamente melhor uma imagem 2D. Os pesos são menores e partilhados entre si, ou seja torna-se mais fácil de treinar o modelo, como existe um menor desperdício dos recursos, ou seja um menor número de parâmetros. Não sendo *fully connected*, mas sim de forma esparsa, permitem ter uma maior profundidade, ou seja maior número de camadas. Relacionando com imagens, as CNN's usam-se do facto de píxeis próximos entre si, estarem mais relacionados entre si, do que píxeis mais distantes. Permitem também encontrar padrões similares, mesmo com imagens que sofreram rotação ou translação, ao contrário das MLP's.

Comprova-se com isto, que, para o processamento de imagens, as CNN's apresentam largas vantagens em relação às MLP's, como é observável pelos accuracy scores, pois para CNN apresenta 0.8962, enquanto para MLP e sem Early Stopping tem 0.8742 e 0.8577, respetivamente.

1.3

Comparando a matriz de confusão obtida a partir do modelo da CNN com a obtida para o modelo da MLP, observa-se um número mais elevado de parâmetros corretamente estimados, pois os valores em cada posição da diagonal da matriz são maiores que para os mesmos na matriz de confusão da MLP, e de igual forma, os valores espalhados pelas outras posições da matriz, que não na diagonal, são no total menores que para a matriz da MLP.

$$C_{MLP} = \begin{bmatrix} 823 & 3 & 14 & 38 & 4 & 2 & 103 & 0 & 13 & 0 \\ 3 & 967 & 1 & 21 & 5 & 0 & 3 & 0 & 0 & 0 \\ 12 & 1 & 816 & 24 & 91 & 0 & 52 & 0 & 4 & 0 \\ 21 & 13 & 9 & 908 & 23 & 1 & 22 & 0 & 3 & 0 \\ 0 & 0 & 104 & 49 & 801 & 0 & 44 & 0 & 2 & 0 \\ 0 & 0 & 1 & 0 & 0 & 949 & 0 & 30 & 1 & 19 \\ 131 & 1 & 95 & 39 & 73 & 0 & 649 & 0 & 12 & 0 \\ 0 & 0 & 0 & 0 & 0 & 22 & 0 & 961 & 1 & 16 \\ 4 & 3 & 7 & 6 & 5 & 1 & 5 & 6 & 961 & 2 \\ 0 & 0 & 0 & 1 & 0 & 5 & 1 & 48 & 0 & 945 \end{bmatrix}$$

$$C_{CNN} = \begin{bmatrix} 870 & 1 & 21 & 18 & 5 & 2 & 76 & 0 & 7 & 0 \\ 2 & 977 & 1 & 14 & 3 & 0 & 3 & 0 & 0 & 0 \\ 12 & 0 & 857 & 11 & 64 & 0 & 55 & 0 & 1 & 0 \\ 22 & 8 & 10 & 886 & 40 & 0 & 32 & 0 & 2 & 0 \\ 1 & 0 & 50 & 26 & 882 & 0 & 39 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 971 & 0 & 18 & 1 & 10 \\ 139 & 1 & 79 & 28 & 99 & 0 & 636 & 0 & 18 & 0 \\ 0 & 0 & 0 & 0 & 0 & 11 & 0 & 956 & 1 & 32 \\ 3 & 1 & 5 & 3 & 2 & 2 & 3 & 5 & 976 & 0 \\ 0 & 0 & 0 & 0 & 0 & 5 & 0 & 35 & 1 & 959 \end{bmatrix}$$

1.4

As imagens de activação representam as várias características detectadas, sendo elas, por exemplo a cor ou as formas. Em camadas posteriores e mais profundas, encontram-se características mais complexas, resultado da combinação das características encontradas anteriormente. Entre camadas, existe a chamada convolução, ou seja, existe um redimensionamento da camada, diminuindo o número de píxeis, verificável na evolução das Figuras 3, 4, 5 e 6, onde se observa uma diminuição do tamanho dos quadrados referente a uma diminuição no número de píxeis utilizados, isto devido a uma das principais características das CNN's de partirem do correcto pressuposto que píxeis próximos se encontram mais fortemente relacionados que píxeis distantes.

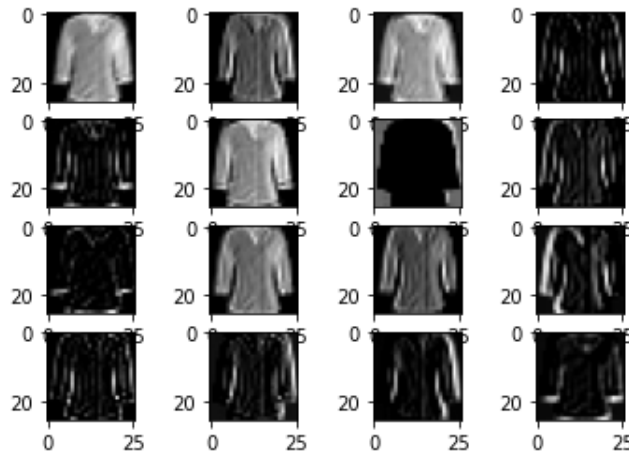


Figura 3: Imagens de activação da primeira camada

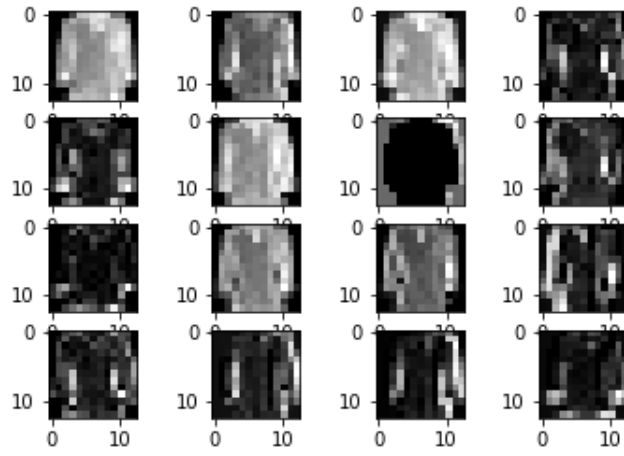


Figura 4: Imagens de ativação da segunda camada

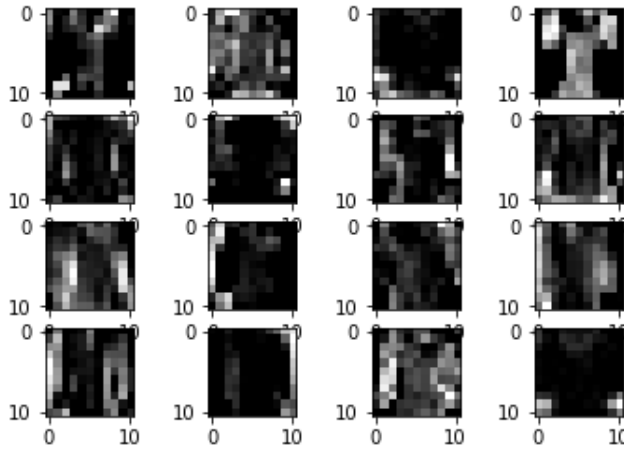


Figura 5: Imagens de ativação da terceira camada

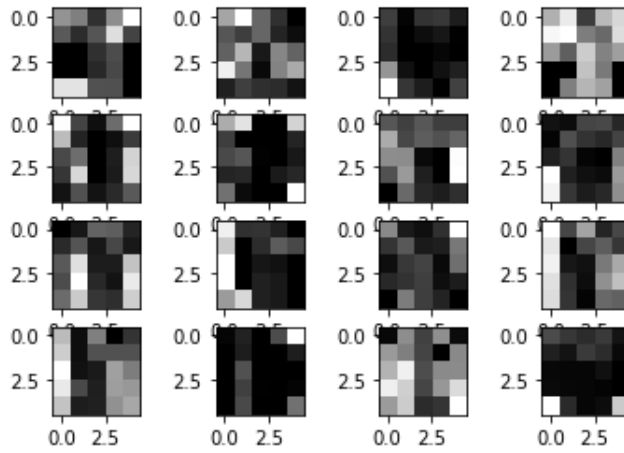


Figura 6: Imagens de ativação da quarta camada