

Prática: Redes Neurais Convolucionais 1 (Deep Learning) (II)

Eduardo Prasniewski

1 Seção 5

Assim como discutido no módulo anterior é explicado de forma mais sucinta neste módulo, abordando uma forma mais matemática, inclusiva a notação da convolução, que trata-se basicamente de uma soma de produtos (neste caso, dos pixels com os kernels). Apresenta a importância da similaridade de cossenos assim como produto escalar. Imagens coloridas possuem 3 canais (RGB), sendo assim a dimensão de entrada é Altura x Largura x 3. Nesta seção eu senti muita falta do que ele explicou em seções anteriores, talvez trocar o módulo anterior pelas seções deste curso seja uma boa opção (já que ele explica mais a fundo as redes neurais, matematicamente).

Também é dedicado a explicar a arquitetura de uma CNN, englobando matematicamente os conceitos de pooling, flatten, Batch Normalization e Augmentation. Foi uma seção muito parecida com o módulo anterior, porém com mais ênfase técnica.

2 Seção 6

A entrada de uma rede neural são números, uma vez que há cálculos em cima dos mesmos, porém no quesito de palavras não é possível apenas jogar o valor numérico da palavra (ASCII), pois são objetos categóricos, sendo assim há um contexto de relação entre as mesmas. Pelo mesmo motivo não é possível usar hot encoding, pois a distância euclidiana entre todas as palavras é a mesma. Assim surge o embedding, que transforma uma palavra em um inteiro e depois em um vetor, e a partir de uma base de dados textual grande treina os pesos e os valores para que palavras similares se situem próximas.

Para fazer isso é necessário tratar o dado de entrada, seja tokenizando as strings e formatando (adicionando padding) para que todas tenham o mesmo tamanho, já que uma rede neural tem um tamanho fixo de entrada.

3 Seção 7

Nesta seção é apresentado os casos de uso de uma Rede Neural Convolucional, como em imagens e áudios.

4 Seção 8

Explica como funciona a arquitetura de uma CNN em imagens 3D (largura, altura e canal RGB), porém parece ser uma seção redundante, visto que já foi explicado de forma minuciosa em seções anteriores.

5 Seção 9

Apresenta a arquitetura base de uma CNN, que apesar de ser simples (camada convolucional seguida pela camada totalmente conectada) é a base dos algoritmos mais famosos e precisos, estes que diferenciam apenas dos hiper-parâmetros usados. Um ponto importante a ser ressaltado é que como são algoritmos que exigem muito do hardware, sendo assim dependem da evolução do mesmo.

6 Seção 10

Primeiramente deduz a equação MSE (Mean Squared Error) e achei muito interessante pois nunca tinha visto desta maneira, apenas a dedução lógica em calculo numérico para a realização da regressão linear. PMF (dados discretos, Binary Cross Entropy), PDF (dados contínuos, MSE), Categorical PMF (Categorical Cross Entropy, dados categóricos). Achei interessante ele passar essa dedução porém eu acho mais didático a forma com que o canal StatQuest no YouTube apresenta, com mais minutagem obviamente por ser um tópico extenso.

7 Seção 11

O gradiente descendente é o coração das redes neurais, é a partir dele que é repassado a direção em que o modelo deve convergir. Possui um hiper-parâmetro chamado taxa de aprendizagem, o qual vai diminuindo de acordo com as iterações pois a cada passo está mais perto do mínimo global da função. SGD (Stochastic Gradient Descent) trata-se de uma forma mais rápida de computar, usando apenas parte da grande base de dados os chamados Batches.

Uma otimização para o gradiente descendente é a adição do momento, fazendo com que converja mais rápido ao valor mínimo global.

A fórmula do gradiente descendente mais usada nos dias atuais é Adam (Adaptive Moment Estimation) que utiliza uma matemática levemente robusta (da qual não consegui entender ao todo por completo, mas compreendi de forma sucinta quase todas as variáveis), normalmente funciona de bom grado para quase todas as aplicações de machine learning.