



MATERIAL COMPLEMENTAR

Deep Learning I

Professor Jeronimo Marcondes Pinto

SUGESTÃO DE LEITURA E VÍDEO

Material disponível na internet:

- Krose, B.; Krose, B.J.A.; Van der Smagt, P.; Smagt, P. 1993. **An introduction to neural networks.** [Link](#)
- I2tutorials. 2019. **Explain Deep Neural network and Shallow neural networks?** [Link](#)
- Rodrigo Guerra. 2021. **Teorema da Aproximação Universal.** [Link](#)

Introdução

Plano de estudo:

O objetivo é compreender as redes neurais multicamadas, identificar o problema de overfitting (quando o modelo se ajusta demais aos dados de treinamento), ajustar os hiperparâmetros (valores que controlam a aprendizagem do modelo) e buscar uma solução ótima que generalize bem para novos dados.

Problemas com a função XOR:

Um desafio inicial das redes neurais foi resolver problemas como a função XOR (Exclusive OR), que não é linearmente separável, limitando a capacidade de redes simples. Isso levou a um período em que o interesse por redes neurais diminuiu até que novos avanços solucionaram essa limitação.

1. Redes Neurais Multicamadas (MLP - Multilayer Perceptron)

- **Deep Learning:**

Redes com múltiplas camadas intermediárias permitem o aprendizado de relações não lineares. A adição de camadas (ou seja, mais profundidade) cria a capacidade de aprendizado profundo, essencial para identificar padrões complexos.

- **Camadas Densas:**

Cada neurônio em uma camada densa conecta-se a todos os neurônios da camada anterior, resultando em uma estrutura totalmente interligada. Isso permite a propagação da informação e a capacidade de capturar relações detalhadas nos dados.

2. Função de Perda e Funções de Erro

- **Função de Perda:**

Essa função mede a precisão da previsão do modelo, comparando a saída prevista com a saída real. É essencial para o aprendizado, pois orienta a atualização dos pesos.

- **Tipos de Funções de Erro:**

- **Erro Binário e Categórico:** Em problemas de classificação, o **binary cross-entropy** é usado para classes binárias (como 0 e 1), enquanto o **categorical cross-entropy** é usado para múltiplas categorias.

- **Erro Quadrático Médio (MSE):** Utilizado para problemas de regressão, mede a diferença média entre os valores previstos e os reais.

3. Viés e Variância

Equilíbrio:

No aprendizado de máquina, busca-se um modelo que tenha baixo **viés** (capacidade de acerto na média) e baixa **variância** (capacidade de generalizar bem para novos dados).

Problemas:

- **Underfitting:** Quando o modelo é simples demais para capturar os padrões dos dados, resultando em uma performance ruim até no conjunto de treinamento.
- **Overfitting:** Quando o modelo se ajusta tão bem aos dados de treinamento que perde a capacidade de generalização para novos dados.

4. Técnicas de Redução de Overfitting

Regularização:

Introduz “barulho” no modelo para reduzir o overfitting. **Dropout**, por exemplo, desativa aleatoriamente neurônios durante o treinamento, impedindo que o modelo memorize os dados de treinamento.

L1 e L2 Regularização:

Ajustam os pesos para evitar que sejam superestimados. Na L1, pesos irrelevantes são reduzidos a zero, enquanto na L2, pesos grandes são reduzidos para valores próximos de zero.

Early Stopping:

Para o treinamento quando a performance no conjunto de validação não melhora mais, evitando o excesso de aprendizado.

5. Épocas e Batch Size

Épocas:

Uma época representa uma passagem completa por todos os dados de treinamento. Mais épocas significam que o modelo ajusta os pesos várias vezes, buscando melhorar a previsão.

Batch Size:

Lotes de dados são processados de uma vez para otimizar a memória e o tempo de cálculo. Em cada lote, os pesos são atualizados para minimizar a função de perda.

6. Hiperparâmetros

Definição:

São parâmetros que controlam o comportamento do treinamento, como **learning rate** (taxa de aprendizado), **batch size** e número de épocas. Esses valores são ajustados com técnicas como **grid search** e **cross-validation**.

Exemplos:

Learning Rate: Determina o tamanho dos passos dados durante a descida do gradiente. Um valor muito alto pode fazer o modelo divergir, enquanto um muito baixo torna o treinamento lento.

7. Otimizadores

Descida do Gradiente e Variantes:

Além da descida do gradiente clássica, outros otimizadores, como **ADAGRAD**, **ADAM** e **RMSPROP**, são projetados para acelerar o aprendizado e melhorar a performance.

Momentum:

Técnica que simula inércia, permitindo que o modelo continue avançando na direção do gradiente mesmo em pontos de “sela” (mínimos locais), ajudando a alcançar o mínimo global.