

Projeto Deep Learning

Advanced Institute for Artificial Intelligence

<https://advancedinstitute.ai>

Agenda

- ☐ Hipótese
- ☐ Preparação de dados
- ☐ Desenho do modelo
 - Underfit x Overfit
 - Estabilidade do modelo por tamanho de lote (batch)
- ☐ Avaliação do modelo
 - Validação Cruzada
- ☐ Utilização do Modelo

Hipótese

- ☐ Definir qual pergunta deve ser respondida pelo modelo
- ☐ A necessidade de negócio deve ser traduzida para um formato de características (features) e anotação (label), quando houver anotação de dados
- ☐ A definição da hipótese está relacionada com a representatividade intrínseca dos dados disponíveis

Exemplo

- ❑ Uma rede de ferrovias e dados históricos de viagem de trens com características diversas: maquinista, carga, origem, destino, data, distância percorrida, tempo de espera por manobras, posição atual, etc
- ❑ Cada linha no dataset corresponde a cada etapa do percursos, e não ao percurso total
- ❑ Utilizar deep learning para criar um dashboard de tráfego de trens, mostrando horários de chegada previsto

Possíveis hipóteses:

- ☐ Predizer o tempo total de viagem
- ☐ Predizer o tempo total por trecho
- ☐ Predizer o atraso por viagem/trecho
- ☐ Modelo de predição por cada tipo de carga ou empresa

Processo exploratório:

- ☐ Não é obrigatório definir APENAS uma
- ☐ Não é obrigatório definir tudo no início
- ☐ A hipótese pode ser melhorada a medida que o modelo evolui

Preparação de dados:

- As features e labels devem ser organizadas em um formato de matriz
 - Todas as amostras devem ter o mesmo número de elementos em todas as dimensões
 - Essa é uma característica difícil de ser obtida naturalmente das fontes de dados
- Definição do formato da entrada e da saída da rede neural
 - O formato da camada de entrada deve ter o mesmo formato da matriz de features
 - O formato da camada de saída deve ter o mesmo formato da matriz de labels
- Limpezas (faltantes e inválidos) e preparações (padronização, normalização) são realizadas antes de submeter os dados ao modelo

Como escolher as features mais relevantes?

- ☐ Manualmente, criar visualizações para comparar as características da amostra com o label
- ☐ Teste estatístico
- ☐ Métodos de redução de dimensionalidade (PCA)

Qual a importância da seleção de features?

- ☐ Diminuir custo de processamento
- ☐ Aumento da capacidade de generalização do modelo
- ☐ Identificar novas oportunidades

Diversos parâmetros podem ser definidos para desenhar um modelo:

- ☐ Número de camadas
- ☐ Parâmetros de camada
 - Neurônio
 - Função de ativação
- ☐ Otimizador
- ☐ Tamanho de lote

Alguns parâmetros são definidas de acordo com a natureza do problema:

☐ Loss

- Mean_square_error para regressão
- Binary_cross_entropy para classificação binária
- categorical_cross_entropy para classificação multilabel

☐ Ativação (Na última camada)

- softmax é mais usada para classificação multilabel
- sigmoid é mais usada para classificação binária
- Relu é mais usada para regressão

☐ Para outros parâmetros é comum utilizar rotinas de busca por hiperparâmetros)

CrITÉrios para definir os parâmetros do modelo:

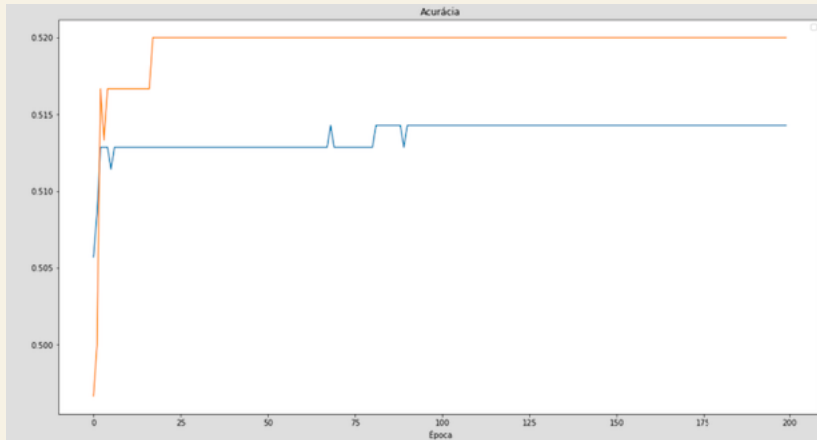
- ☐ Balanço overfit x underfit
 - Curva de loss
- ☐ Estabilidade do modelo
 - Variações na Curva de loss

Estabilidade do modelo (definições):

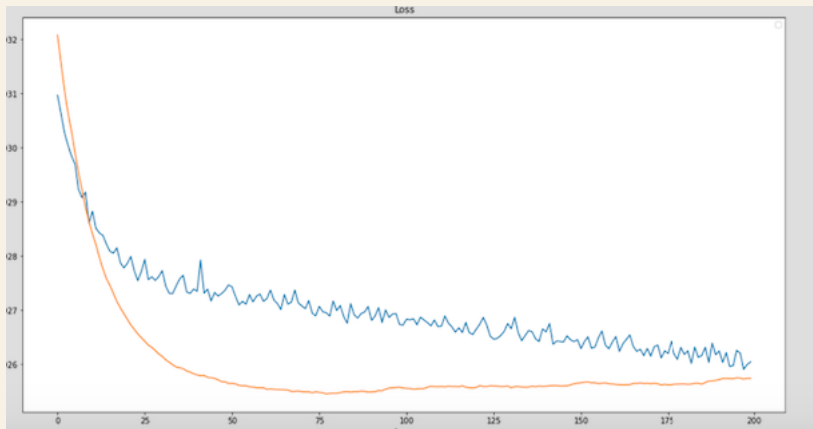
- ❑ Amostra: um elemento de um conjunto de dados. Exemplos: uma imagem, um arquivo de áudio, uma linha no banco de dados
- ❑ Lote (batch): um conjunto de N amostras. As amostras em um lote são processadas independentemente, em paralelo. Durante o treino um lote resultará em apenas uma atualização para o modelo
- ❑ O tamanho do lote influencia: a ocupação da memória durante o treino, a quantidade de atualizações do modelo e a estabilidade do modelo
- ❑ Lotes grandes ocupam muita memória, o que pode tornar o treino inviável ou lento de acordo com o hardware, lotes pequenos podem levar a muitas atualizações de modelo, tornando lento o treinamento

Projeto Deep Learning

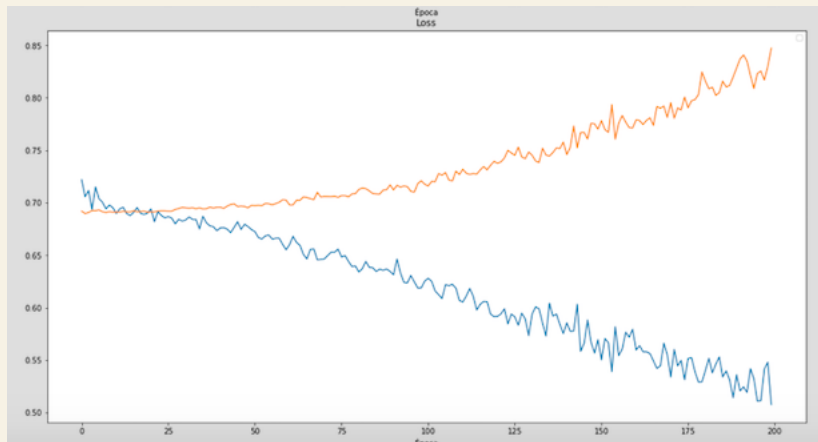
Acurácia em Underfit



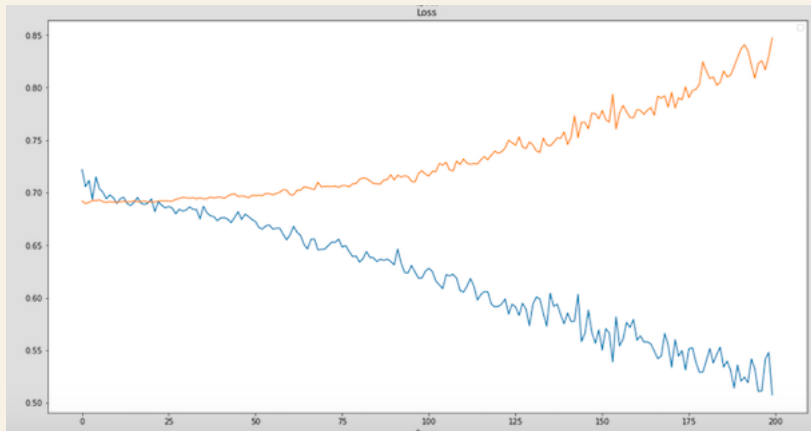
Loss em Underfit



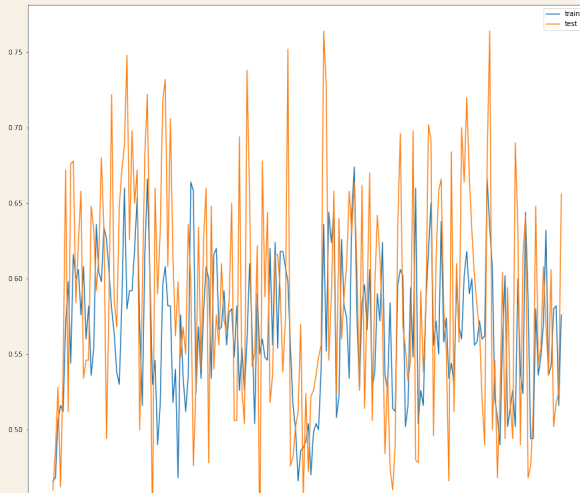
Acurácia em Overfit



Loss em Overfit

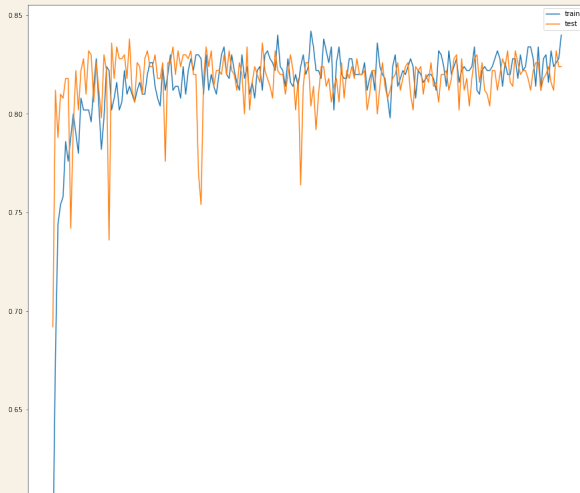


Modelo em Underfit sem estabilidade



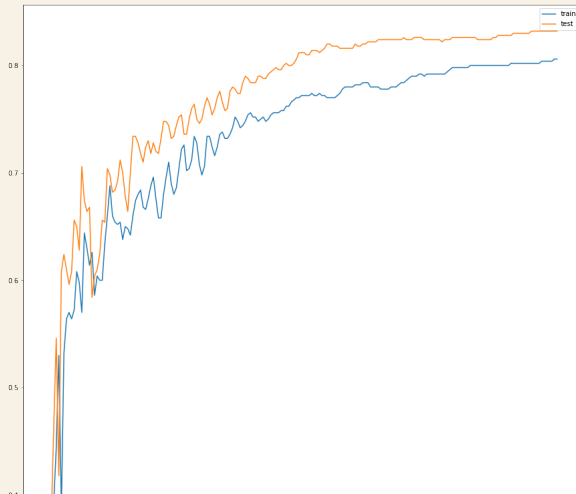
Projeto Deep Learning

Modelo com boa acurácia mas sem estabilidade



Projeto Deep Learning

Modelo com boa acurácia e estabilidade



- O tamanho de lote pode ser ajustado para encontrar um bom balanço entre overfit x underfit e estabilidade do modelo
- Otimizador pode levar em conta além do valor do peso, a evolução dos pesos
- Algumas modelos adaptam os parâmetros do otimizador, de acordo com o loss a cada passo iterativo
 - Ajuste dinâmico do learning rate
 - Parada antecipada, quando o loss se altera menos do que um limite pré-estabelecido

Avaliação do modelo:

- ❑ A avaliação do modelo é feita dividindo os dados em treino e teste
- ❑ A avaliação de um modelo de deep learning é feita em cada etapa do treino, pois é um método iterativo
 - Nessa etapa o conjunto de treino é dividido em treino e validação
- ❑ Após o treino do modelo, é necessário avaliar com um conjunto desconhecido para identificar o desempenho do modelo de modo confiável
 - Nessa etapa é utilizado o conjunto de teste, desconhecido do modelo durante todo o treino

- Com o modelo treinado é possível utilizar todas as técnicas clássicas: matriz confusão , LOOCV, kfold, etc
- Para a comunicação mais clara dos resultados para não-especialistas em Deep-learning é recomendável criar critérios de qualidade próprio
- Exemplos:
 - Quantidade de imagens processados por intervalo de tempo
 - Porcetagem de acerto por áreas de interesse (não necessariamente por classes do problema)

Colocando em produção:

- ☐ Depois que o processo de treino termina e os resultados são satisfatórios, o binário do modelo deve ser usado para inferências diretamente
- ☐ O modelo pode ser constantemente retreinado, e adaptar-se de acordo com a dinâmica do problema
- ☐ Parte do modelo pode ser reutiliza em outros contextos (transfer learning)