

TÓPICOS EM CIÊNCIA DE DADOS PARA O ESPORTE

MACHINE LEARNING: CLASSIFICAÇÃO

DIEGO RODRIGUES DSC

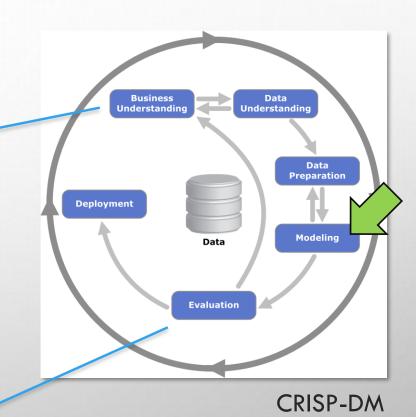
INFNET

AGENDA

- PARTE 1 : TEORIA
 - CONCEITOS
 - REDES NEURAIS
 - TREINANDO UMA REDE NEURAL
 - CASE : CLASSIFICAÇÃO DE JOGADAS DE FUTEBOL AMERICANO

CRONOGRAMA

DIA	NÚMERO	ÁREA	AULA	TRABALHOS
10/10/2023	1	Intro	Introdução a Disciplina e Organização do Ambiente	
17/10/2023	2	Dados	Coleta de Dados e Sensoriamento	
19/10/2023	3		Variáveis Aleatórias	Grupos
24/10/2023	4		Análise Exploratória	
26/10/2023	5		Estatísticas para Ranqueamento	
07/11/2023	6	Estatística	Ranqueamento Estatístico : ELO	
09/11/2023	7		Ranqueamento Estatístico : Glicko	
14/11/2023	8		Ranqueamento Estatístico : TrueSkill	
16/11/2023	9		Ranqueamento Estatístico : XELO	Base de Dados
21/11/2023	10		Modelos de Aprendizado de Máquina	V
23/11/2023	11		Machine Learning: Classificação	
28/11/2023	12	ML	Machine Learning: Regressão	Pesquisa
30/11/2023	13		Machine Learning: Agrupamento	
5/12/2023	14		Machine Learning: Visão Computacional	
7/12/2023	15		Aplicações & Artigos: Esportes Independentes	
12/12/2023	16	Esportes	Aplicações & Artigos: Esportes de Combate	Modelo
14/12/2023	17	Laportes	Aplicações & Artigos: Esportes de Objeto	
19/12/2023	18		Aplicações & Artigos : Betting	
21/12/2023	19	Workshop	Workshop	
4/1/2024	20	worksnop	Apresentações de Trabalhos	Apresentação



SETUP INICIAL DO AMBIENTE PYTHON



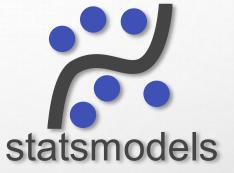
4. Variáveis Aleatórias



5. Visualização

6. Estimação e





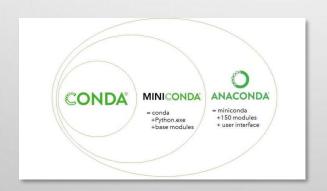


7. Machine Learning





1. Editor de Código



2. Gestor de Ambiente



3. Ambiente
Python do Projeto

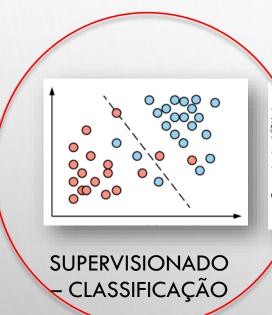


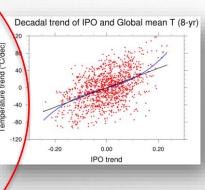
3. Notebook Dinâmico

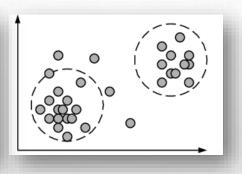


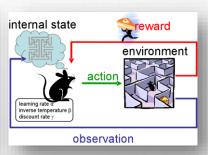
CONCEITOS

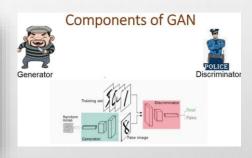
PARADIGMAS DE MODELAGEM ESTATÍSTICA











SUPERVISIONADO – REGRESSÃO NÃO SUPERVISIONADO

APRENDIZADO POR REFORÇO **GENERATIVO**



PROBLEMA DE NEGÓCIO

Classificação

Um bebê consegue separar e ordenar blocos com diferentes tamanhos, formas e cores. Ele também consegue identificar os tipos diferentes de objetos.

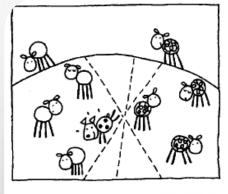
Os diferentes tipos de objetos são chamados de **classes**. As características dos objetos são chamadas de **variáveis** ou **atributos**.

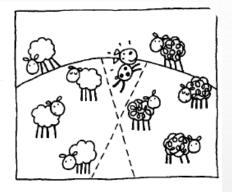


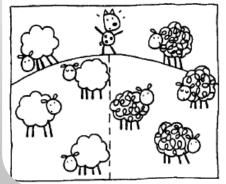
Então, um classificador é um modelo treinado para discriminar objetos pertencentes a duas ou mais classes, baseado em seus atributos.

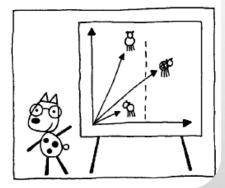


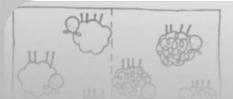
REPRESENTAÇÃO













Exercício (1): qual representação o cachorro deve escolher para diferenciar ovelhas pretas, cinzas e brancas?

Exercício (2): qual seria uma boa representação para diferenciar alunos e alunas do curso?



MODELAGEM

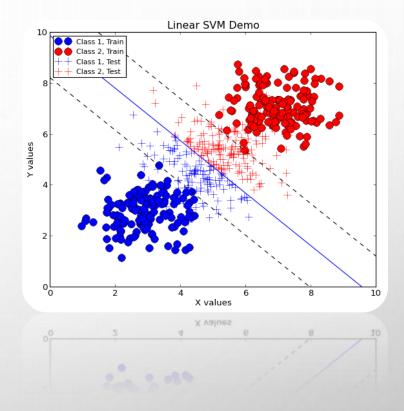
1) Aprendizado Supervisionado

Tarefas de classificação e regressão pertencem a esta categoria. O treinamento consiste em encontrar parâmetros para o modelo que minimiza uma função de risco/erro para uma amostra de treinamento, baseado na diferença entre os valores previstos e reais, para cada observação.

MODELOS FUNCIONAIS

Algoritmos que dependem da estimação dos parâmetros de uma função que é utilizada como superfície de separação entre as classes.

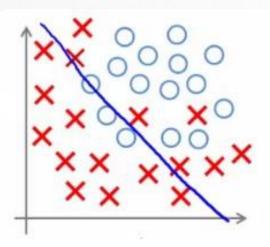
- 1) Funções Polinomiais
- 2) Regressão Logística
- 3) Máquina de Vetores Suporte
- 4) Neurônio Sigmoide / Tangente Hiperbólica
- 5) Árvores de Decisão



Algoritmos baseados em funções são **mais simples**, usualmente tem um **número menor de parâmetros** e não dependem em armazenar muitos dados para manter uma "memória", como por exemplo K-vizinhos mais próximos.



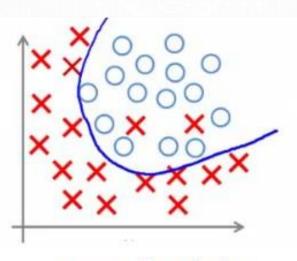
VALIDAÇÃO



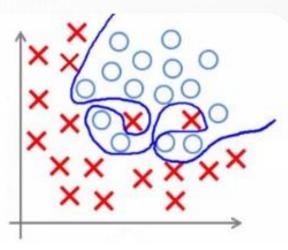
Under-fitting

(too simple to explain the variance)

explain the variance)



Appropriate-fitting



Over-fitting

(forcefitting -- too good to be true)





GENERALIZAÇÃO: IDENTIFICANDO OS HIPERPARÂMETROS ÓTIMOS

LEAVE ONE OUT

 Uma única observação é deixada de fora a cada treinamento. N treinamentos são realizados para calcular a estatística de erro.

K FOLDS

 Amostra é dividida em K conjuntos. K treinamentos são realizados, mantendo um conjunto como fora-da-amostra.

BOOTSTRAPPING

 O algoritmo itera, amostrando aleatoriamente M observações, para a quantidade Q desejada de treinamentos.

relevant elements false negatives true negatives 0 true positives false positives selected elements How many selected How many relevant items are relevant? items are selected? Recall = -Precision = -

FIGURAS DE MÉRITO

Acurácia

• (TP+TN)/(P+N)

Taxa de Erro

1-Acurácia

Sensibilidade (Recall)

TP/(TP+FN)

Especificidade

TN/(TN+FP)

Precisão

TP/(TP+FP)

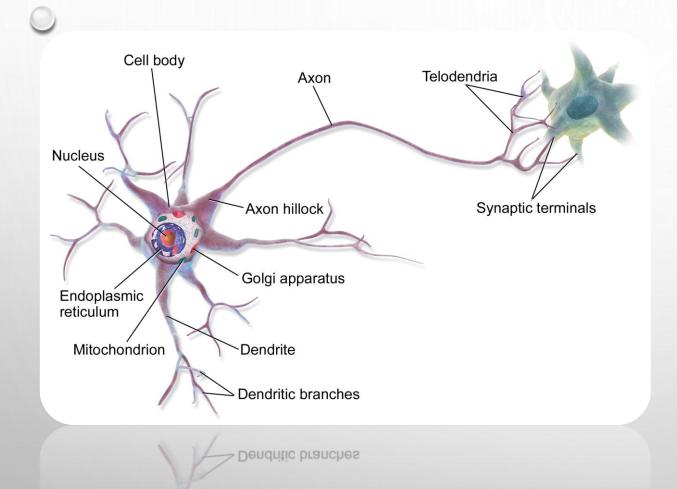
Produto Sp

SQRT[SQRT(R1*R2)

$$*(R1 + R2)/2$$



REDES NEURAIS



INSPIRAÇÃO BIOLÓGICA



HISTÓRIAS DAS REDES NEURAIS

Perceptron de Rosenblatt 1954

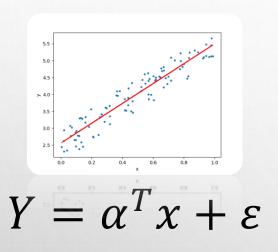
Backpropagation Rumelhart, Hinton & Williams 1986

LSTM ~ 2003

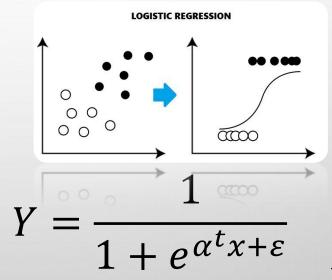
Tensorflow ~ 2015

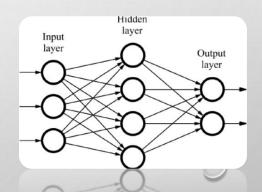
O APROXIMADOR UNIVERSAL

$$Y = F(X) + \varepsilon$$

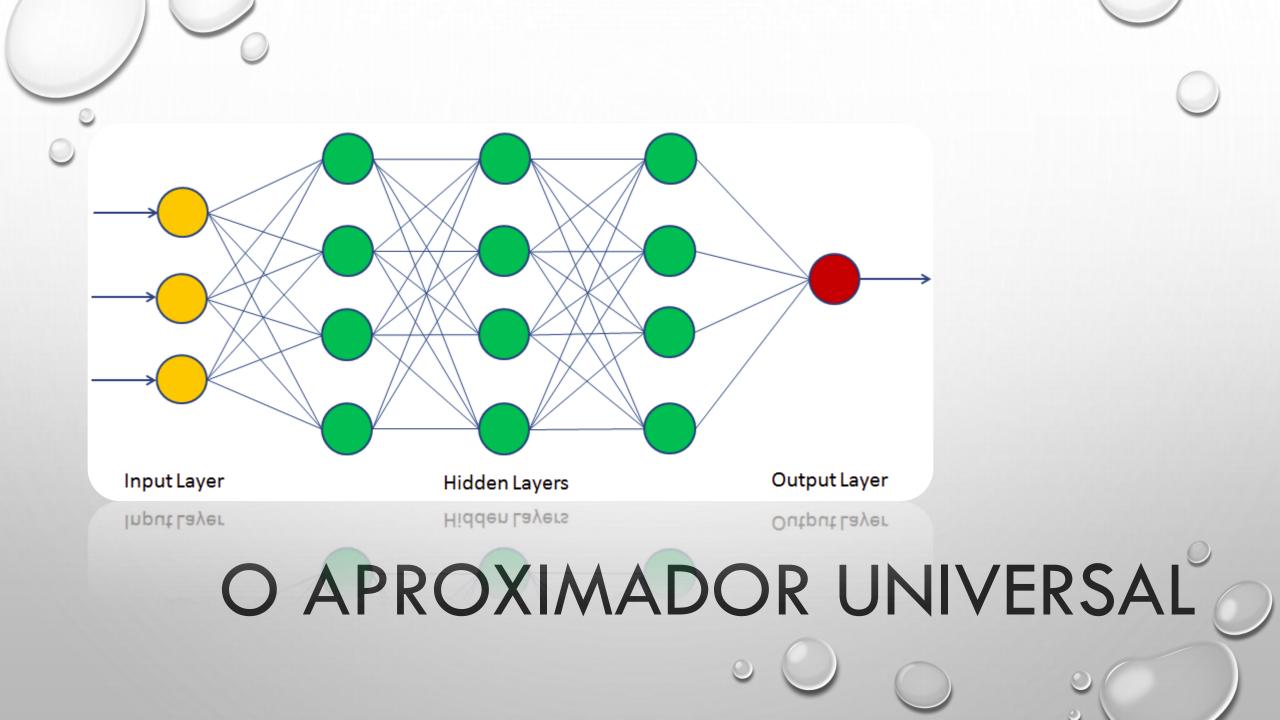


$$Y = X\alpha + \varepsilon$$





$$Y = \varphi(x) + \varepsilon$$



Activation function	Equation	Example	1D Graph
Unit step (Heaviside)	$\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Sign (Signum)	$\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$	Perceptron variant	
Linear	$\phi(z) = z$	Adaline, linear regression	
Piece-wise linear	$\phi(z) = \begin{cases} 1, & z \ge \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \le -\frac{1}{2}, \end{cases}$	Support vector machine	
Logistic (sigmoid)	$\phi(z) = \frac{1}{1 + e^{-z}}$	Logistic regression, Multi-layer NN	
Hyperbolic tangent	$\phi(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$	Multi-layer Neural Networks	
Rectifier, ReLU (Rectified Linear Unit)	$\phi(z) = \max(0, z)$	Multi-layer Neural Networks	
Rectifier, softplus Copyright © Sebastian Raschka 2016 (http://sebastianraschka.com)	$\phi(z) = \ln(1 + e^z)$	Multi-layer Neural Networks	
Rectifier, softplus Copyright © Sebastian Raschka 2016 (http://sebastianraschka.com)	$\phi(z) = \ln(1 + e^z)$	Multi-layer Neural Networks	-
Unit)		Networks	

FUNÇÕES DE ATIVAÇÃO



GOOGLE TENSORFLOW PLAYGROUND



TREINANDO UMA REDE NEURAL



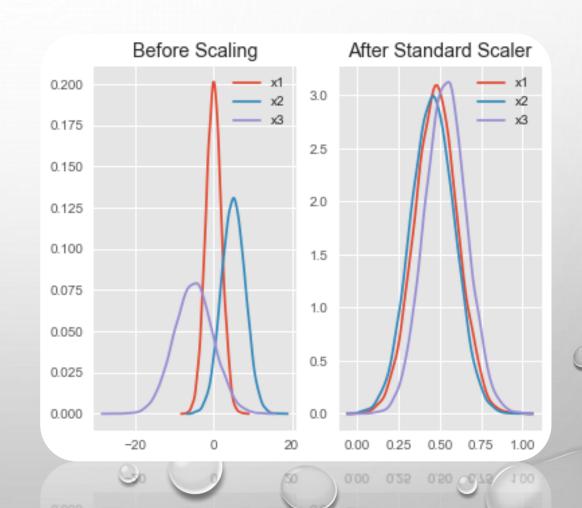
REPRESENTAÇÃO

- Normalização
 - · Garantir que as variáveis independentes possuem a mesma escala
- Análise de Componentes Principais
 - Garantir que as variáveis independentes sejam descorrelacionadas



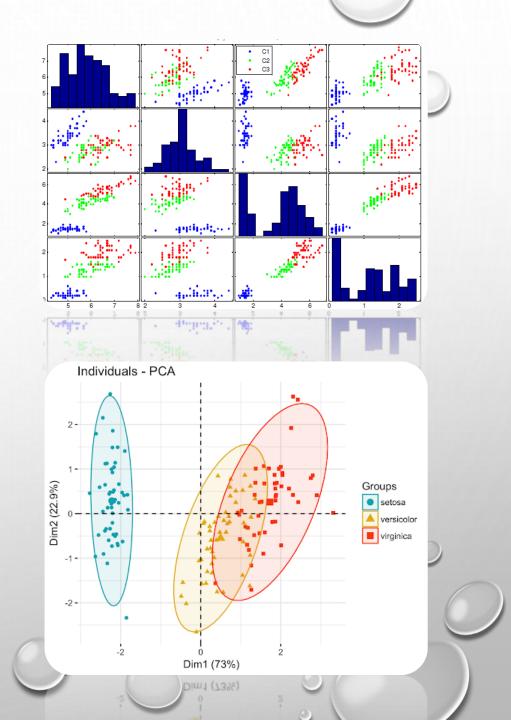
NORMALIZAÇÃO

- Garantir que as variáveis independentes possuem a mesma escala
- Mesmo efeito numérico na otimização independente da escala.



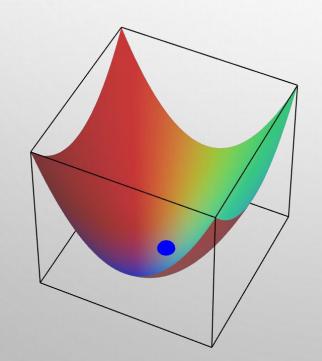
ANÁLISE DE COMPONENTES PRINCIPAIS

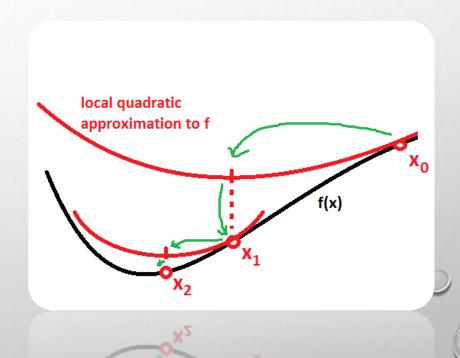
- Garantir que as variáveis independentes sejam descorrelacionadas.
- Identificar novas direções com maior concentração de energia / informação.



ENCONTRANDO O ÓTIMO GLOBAL PARA OS PARÂMETROS

$$MSE = \frac{1}{N} \sum_{i} (y - f(w, x))^2$$



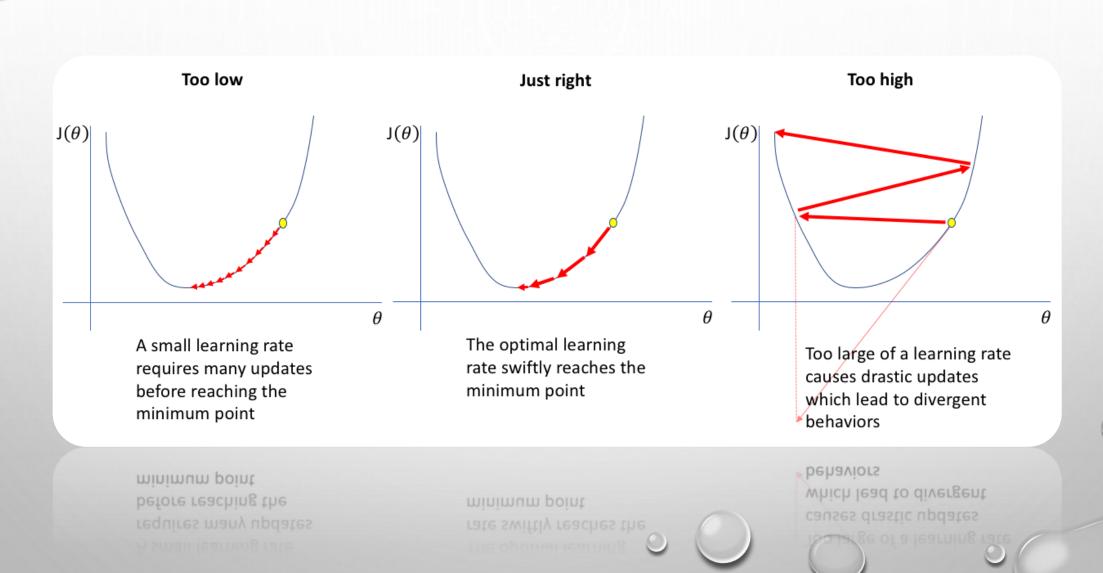


Wolfram Global Problem

SUPERFÍCIE DO ERRO MÉDIO QUADRÁTICO EM FUNÇÃO DOS PARÂMETROS

$$MSE = \frac{1}{N} \sum (y - f(w, x))^2$$

ESTIMADOR DE GRADIENTE + TAXA CORRETA





- TREINAMENTO UTILIZANDO 5 PARTIÇÕES, COM DADOS DAS CLASSES BALANCEADOS.
- CADA TREINAMENTO É REALIZADO PARA EXPLORAR UMA
 CONFIGURAÇÃO DA CAMADA OCULTA DA REDE O NÚMERO
 DE NEURÔNIOS DA CAMADA OCULTA.
- 4 PARTIÇÕES SÃO USADAS PARA TREINAR A REDE, 1 PARTIÇÃO É UTILIZADA PARA MENSURAR A ACURÁCIA FORA DA AMOSTRA.
- A ACURÁCIA MÉDIA É USADA COMO FIGURA DE MÉRITO PARA CADA CONFIGURAÇÃO.

Validação & Teste

Treino

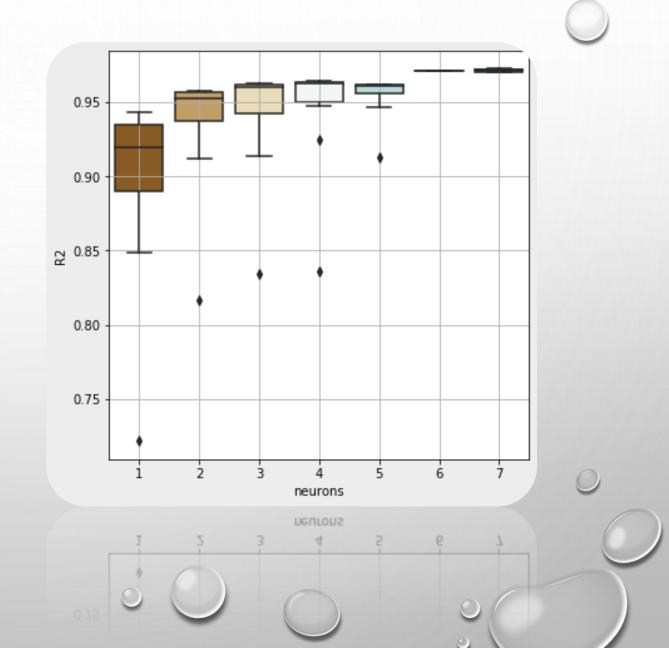
Treino

Treino

Treino



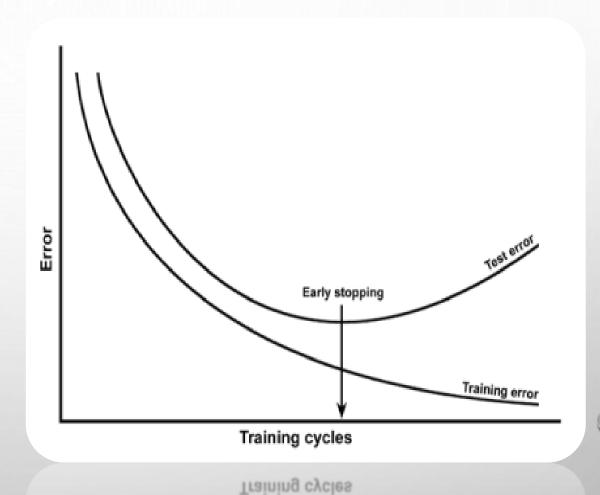
- Iteração dos hiperparâmetros
- Seleção da Figura de Mérito
- Seleção da Estatística de Ganho





CRITÉRIOS DE PARADA

- Aumento no Erro de Validação
- Estabilidade da Figura de Mérito no Treino
- Estabilidade de Figura de mérito no Teste

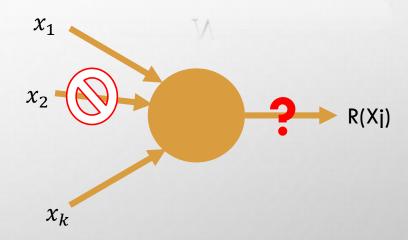




RELEVÂNCIA

- ANÁLISE NÃO-LINEAR DA IMPORTÂNCIA DOS ATRIBUTOS.
- ELIMINA-SE UM ATRIBUTO DA ENTRADA DA REDE, MEDINDO O EFEITO NA CAPACIDADE PREDITIVA.
- A COMPARAÇÃO DAS RELEVÂNCIAS NORMALIZADAS PERMITE IDENTIFICAR OS ATRIBUTOS MAIS IMPORTANTES.

$$R(X_j) = \frac{\sum_{i=1}^{N} \left| |\hat{y}(\mathbf{x_i}) - \hat{y}(\mathbf{x_i}|_{x_{ij} = \bar{x}_j}) \right| |^2}{N}$$

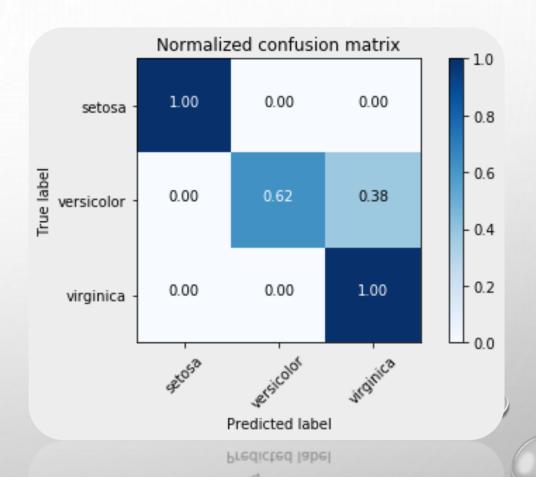


A redução na quantidade de atributos permite simplificar o modelo, salientando os atributos direcionadores da classe.



Matriz de Confusão

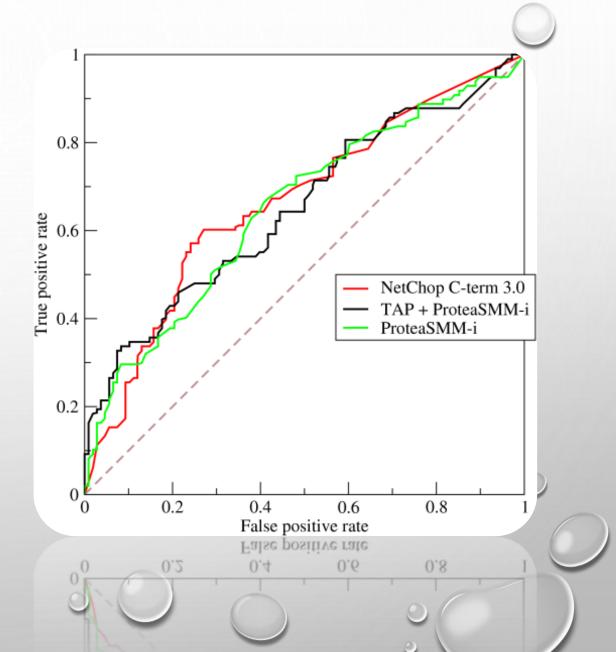
 Comparação entre o resultado do classificador para as diferentes classes.



VALIDAÇÃO

Curva ROC

Calibra a saída do modelo,
 ajudando a configurar o ponto de operação entre Precisão / Recall / Acurácia.





CLASSIFICAÇÃO DE JOGADAS DE FUTEBOL AMERICANO



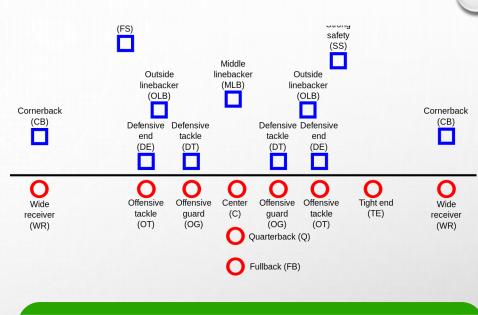
FUTEBOL AMERICANO

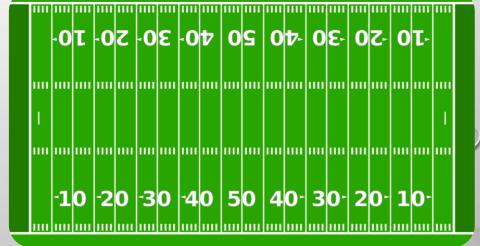
Campo com 54x120 jardas.

Partida disputada em 2 tempos.

Cada tempo é dividido em dois quartos de 15 minutos.

Objetivo: conquistar território e chegar a endzone adversária, para marcar um touchdown ou field goal.





ANÁLISE DE ATRIBUTOS DIRECIONADORES

Representação

- Dados da temporada de 2016 do Indianapolis Colts.
- Quarterback Payton Manning
- Atributos extraídos do Play-by-Play.

Treinamento

- Ensemble de Redes Neurais Multicamadas.
- Treinamento Jacknife.
- Análise de Relevância
- Figura de Mérito: Acurácia



ATRIBUTOS COLETADOS DO PLAY-BY-PLAY

N	Atributo	Tipo
1	Distância para o First Down	Contínuo
2	Posição no Campo	Contínuo
3	Diferença no Placar	Contínuo
4	Quarto de Jogo	Discreto
5	Tempo	Contínuo
6	Dois Minutos Finais	Binário
7	Última Jogada - Passe	Binário
8	Última Jogada - Corrida	Binário
9	Última Jogada - Times Especiais	Binário
10	Última Jogada - Jardas	Contínuo
11	Tipo de Joagada (Corrida/Passe)	Classe

11	Tipo de Joagada (Corrida/Passe)	Classe
10	Última Jogada - Jardas	Contínuo
	Ultima Jogada - Times Espe⊕is	(Bi) SIIO



- REDE NEURAL DO TIPO MULTILAYER PERCEPTRON (MLP) COM UMA CAMADA ESCONDIDA.
 - UMA REDE FOI TREINADA PARA OS DADOS DE CADA DOWN (COMITÊ).
- TREINAMENTO USANDO O ALGORITMO RPROP.
- TREINAMENTO JACKNIFE USANDO ACURÁCIA COMO FIGURA DE MÉRITO – VALIDAÇÃO CRUZADA COM 5 PARTIÇÕES.
 - FASE 1 EXTRAÇÃO DOS DIRECIONADORES.
 - FASE 2 VALIDAÇÃO.



RESULTADOS

REDES NEURAIS: TREINAMENTO FASE 1

Tentativa	Passe	Corrida
Primeiro down	113	135
Segundo down	100	86
Terceiro down	85	12
Total	298	233

Total	298	233

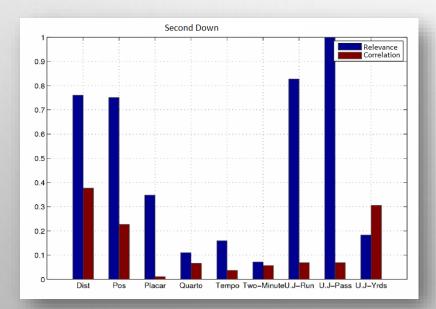
Tentativa	NE	μ (%)	$\sigma(\%)$	Melhor(%)
10 down	20	61	6	71
20 down	9	67	6	78
3o down	8	94	5	97



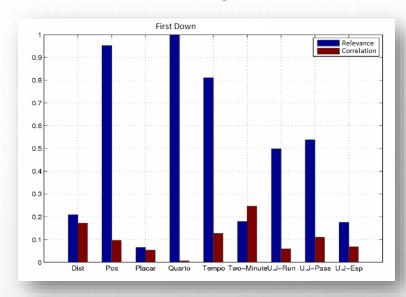
RESULTADOS DA RELEVÂNCIA

Atributo	1D	2D	3D
Distância	21	76	52
Posição	95	75	98
Placar	6	34	41
Quarto	100	11	100
Tempo	81	15	63
2 Minutos	17	7	17
U.J. Corrida	49	82	4
U.J. Passe	53	100	31
U.J. Especial	17	-	-
U.J. Jardas	-	18	10

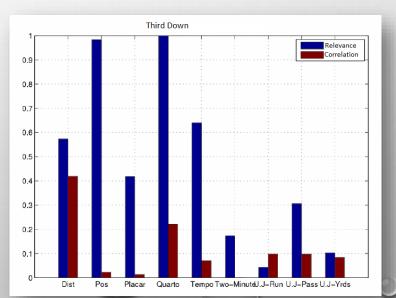
U.J. Passe, U.J. Corrida, Distância



Quarto, Posição, Tempo



Quarto, Posição, Tempo, Distância

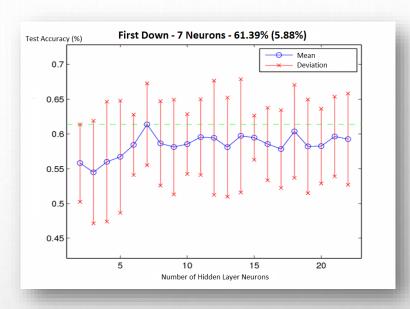


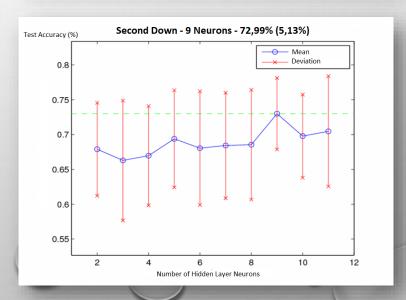
TREINAMENTO 2 – RESULTADO FINAL

Modelo	NE	μ (%)	$\sigma(\%)$	Melhor (%)
NN-1D	7	61	5	74
NN-2D	9	73	5	81
NN-3D	8	94	5	97
NN	10	73	-	73
See5	-	69	-	69
ID3	-	64	-	64

5/10 atributos selecionados como direcionadores para o primeiro down e segundo down.

Eficiência dos direcionadores selecionados confirmado por 80% de acurácia.





DESAFIO: ANÁLISE EXPLORATÓRIA DOS DADOS DO P. MANNING

PRÓXIMA AULA LEITURA: REGRESSÃO