Universidade Federal do Rio Grande do Norte
-
DIMAP

Análise de Dados
e
Interpretação de Resultados

Antonio Carlos Gay Thomé

Problemas e Modelos

No mundo real são inúmeros os problemas cuja solução não é exata e cuja modelagem computacional é muito complexa ou mesmo impossível de ser conseguida por meio de algoritmos convencionais.

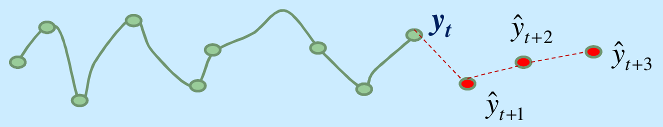
Tais problemas se enquadram nas seguintes classes:

(*) $\left\{ \begin{array}{l} \text{➢ Previsão} \\ \text{➢ Classificação} \\ \text{➢ Reconhecimento ou} \\ \text{➢ Agrupamento} \end{array} \right.$

- ➢ **Previsão**
- ➢ **Classificação**
- ➢ **Reconhecimento ou**
- ➢ **Agrupamento**
- ➢ **Detecção (Localização)**

Previsão

Na previsão a procura é por um modelo que consiga prever valores futuros de uma série de observações ordenadas temporalmente.

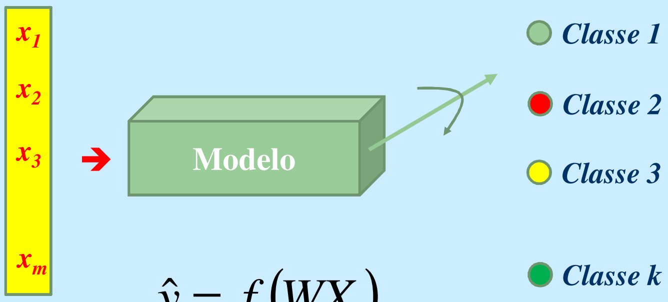


$$\hat{y}_{t+k} = f\left(y_t, y_{t-a}, y_{t-b} \ldots, y_{t-m}\right)$$

Contexto: Univariado x Multivariado
Horizonte: Simples x Múltiplo

Classificação

Na classificação a busca é por um modelo que consiga mapear valores fornecidos em um conjunto finito e pré-definido de classes.

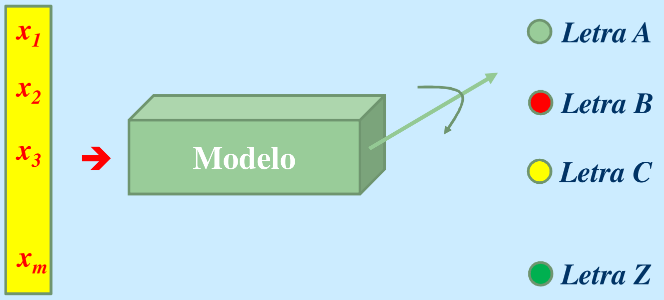


$$\hat{y} = f\left(WX\right)$$

Modelos de Deep Learning

Reconhecimento

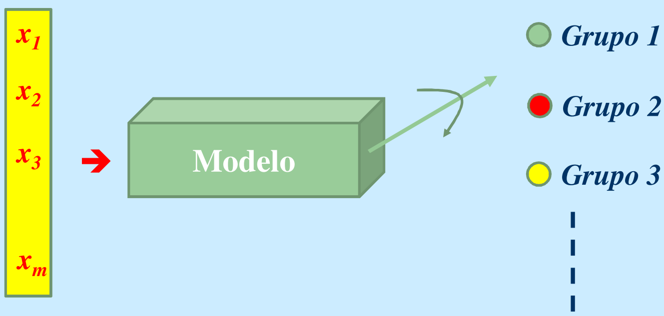
O reconhecimento é um caso particular da classificação, onde cada classe representa uma e somente uma entidade.



Modelos de Deep Learning

Agrupamento

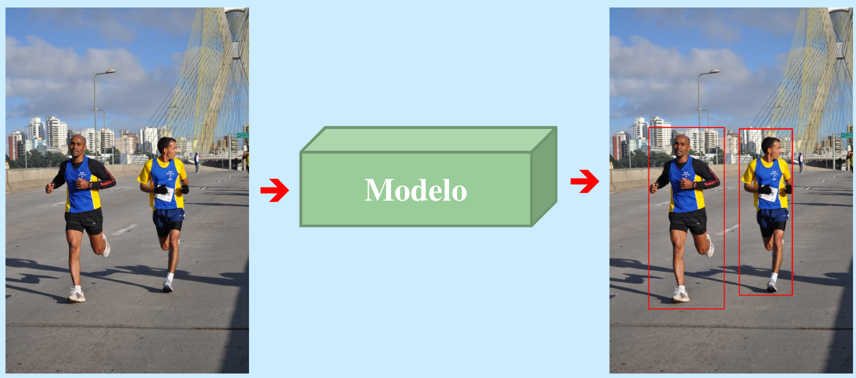
No agrupamento a busca é por similaridades entre as observações de tal forma que permita definir um conjunto finito de grupos que as contenha e as descreva.



(*) Não há um conhecimento prévio quanto ao possível número de grupos

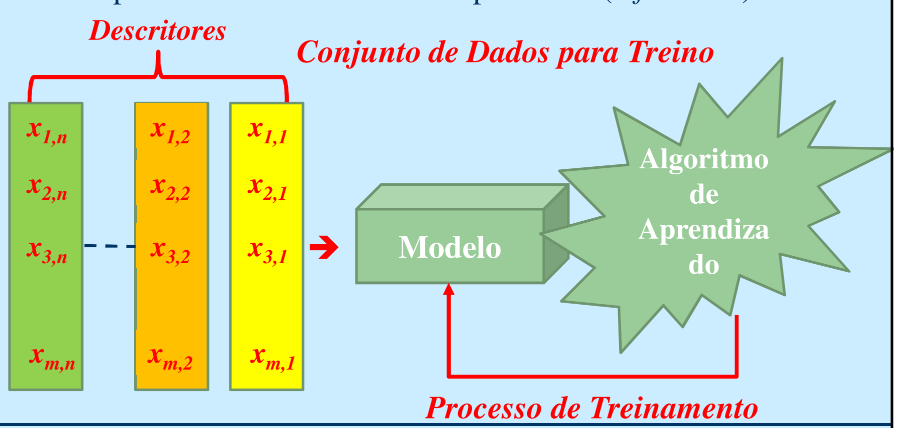
Detecção

Na detecção a busca é por um modelo que seja capaz de localizar os objetos de interesse dentro do contexto do problema.



Modelagem

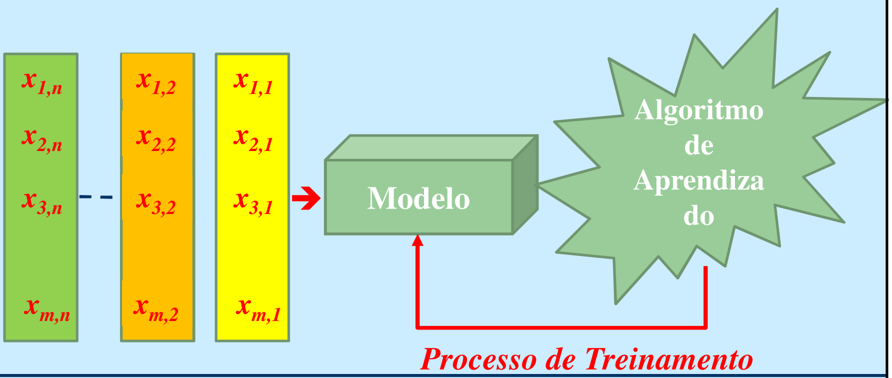
A modelagem de tais tipos de problema é hoje feita através do emprego de técnicas de ***Aprendizagem de Máquina*** onde, o próprio modelo, a partir de exemplos previamente fornecidos, aprende sozinho a resolver o problema (*inferência*).



Modelos de Deep Learning

Modelagem

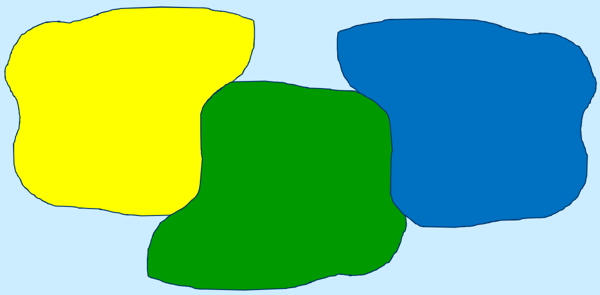
O processo de treino usado nos métodos de aprendizado por inferência busca a solução através de aproximações sucessivas (convergência) e, quanto mais parâmetros livres tiver o modelo, mais dados para treino se fazem necessários.



Modelos de Deep Learning

Seleção da Massa de Dados

O processo de construção de um modelo requer a seleção de três conjuntos de dados: Treino, Validação e Teste.



Condições Ideias:

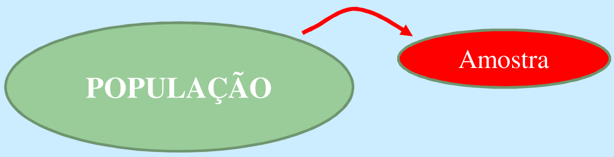
- *Serem igualmente representativos*
- *Serem disjuntos*
- *Terem tamanho suficiente*

População x Amostra

Se o conjunto de dados é *exaustivo*, isto é, inclui toda e qualquer entidade que faça parte do problema, dizemos simplesmente que ele forma uma ***POPULAÇÃO***.

Quando algumas, mas não todas, as entidades estão presentes no conjunto de dados, ele passa a ser uma ***AMOSTRA***.

Amostra é um subconjunto da **População**.



Amostragem

Amostragem é o processo pelo qual os dados são coletados. Isto nos dá apenas uma imagem da população em estudo.

Independentemente da precisão e da correção dos processos de amostragem usados, há sempre a considerar o chamado ***erro de amostragem***.

Devemos sempre esperar algumas diferenças entre a **amostra** e a **população**.

Representatividade

Representatividade é um conceito relativo ao grau de similaridade estatística entre a População de tamanho N e uma Amostra de tamanho n, sendo **n << N**.

A métrica de similaridade estatística usada é tomada com base no cálculo dos momentos estatísticos, geralmente os de 1ª e 2ª ordem, respectivamente, a média e a dispersão em relação à média (variância).

POPULAÇÃO - A

Amostra - B

$$\boxed{\begin{array}{c}\mu_A \approx \mu_B \\ \sigma_A^2 \approx \sigma_B^2\end{array}}$$

Balanceamento

Balanceamento tem a ver com a representatividade da Amostra no tocante às Classes existentes na População e o número de observações de cada Classe usada no ***Treino do Modelo***.

Classe A – frequência de 60%
Classe B – frequência de 30%
Classe C – frequência de 10%

POPULAÇÃO

A – 58%
B – 31%
C – 11%

AMOSTRA

O balanceamento depende do objetivo pretendido no treino

A – 33%
B – 33%
C – 34%

Cj TREINO

(*) O que fazer quando o volume de dados é pequeno e / ou desbalanceado?

Validação Cruzada

Validação cruzada é o particionamento do conjunto de dados em subconjuntos mutualmente exclusivos e seu emprego para estimação do modelo (treino) e posterior validação (teste).

Diversas formas de realizar o particionamento são sugeridas na literatura, sendo as três mais utilizadas:

- O *holdout*,
- O *k-fold* e
- O *leave-one-out*

Validação Cruzada

Holdout

Consiste em dividir o total de dados em dois subconjuntos disjuntos, um para treino (estimação dos parâmetros) e outro para teste (validação). Uma proporção muito comum é 2/3 para treino e 1/3 para teste.

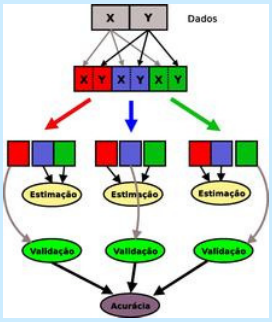
É indicada quando tem-se uma grande quantidade de dados. Caso o conjunto de dados seja pequeno, o erro calculado na validação pode sofrer muita variação.

Validação Cruzada

K-fold

Consiste em dividir o conjunto total de dados em ***k*** subconjuntos ***mutuamente exclusivos do mesmo tamanho*** e, a partir disto, um subconjunto é utilizado para teste e os *k-1* restantes são utilizados para estimação dos parâmetros.

Este processo é realizado *k* vezes alternando de forma circular o subconjunto de teste.



Exemplo com
K = 3

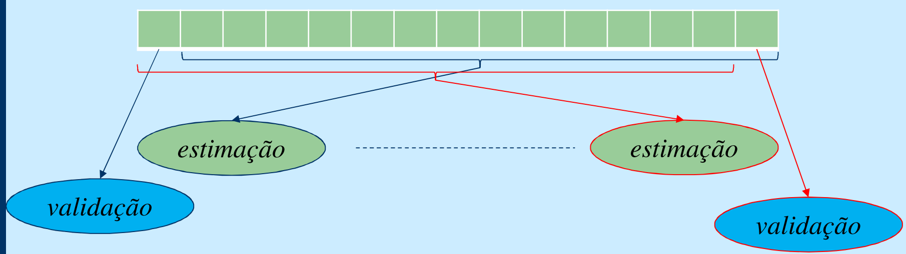
Validação Cruzada

Leave-one-Out

É um caso específico do *k-fold*, com **k igual ao número total de dados N**.

Nesta abordagem são realizados *N* cálculos de erro, um para cada dado.

Este método possui um alto custo computacional, sendo indicado para situações onde poucos dados estão disponíveis.



Emprego dos Conjuntos

Treino:
- Representatividade e Balanceamento são fundamentais
- Volume é crucial para o aprendizado e generalização
- Volume x Ciclos de Treino

Validação:
- Representatividade e Balanceamento são fundamentais
- Importante como Critério de Parada (*overtraining*)

Teste:
- Representatividade e Balanceamento são importantes
- Volume é importante para validação estatística
- Idealmente disjunto dos demais Conjuntos

Interpretação dos Resultados e Análise de Desempenho

Testes para levantamento de desempenho e a análise dos resultados alcançados é um processo que inclui um razoável grau de incerteza.

Uma boa prática no sentido de reduzir o grau de incerteza inclui:

- Selecionar um conjunto de teste representativo e de tamanho estatisticamente satisfatório;
- Dentro do possível fazer uso de mais de um conjunto de testes, idealmente 30;
- Fazer uso de métricas como intervalo de confiança e testes de hipótese para validar os resultados alcançados.

Modelos de Deep Learning

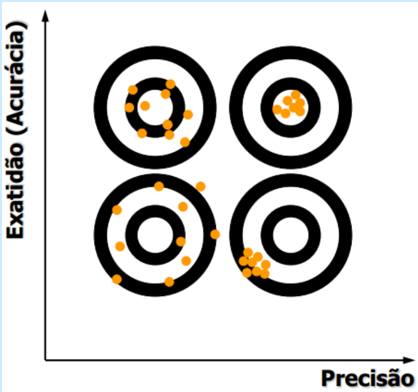
Acurácia x Precisão

Acurácia (exatidão):

É dada pelo grau de concordância entre o *resultado gerado* (estimado) pelo modelo e o *valor verdadeiro* (desejado) do resultado (***Target***).

Precisão:

É o grau de concordância (proximidade) entre vários resultados gerados pelo modelo sob as mesmas condições. (repetitividade)



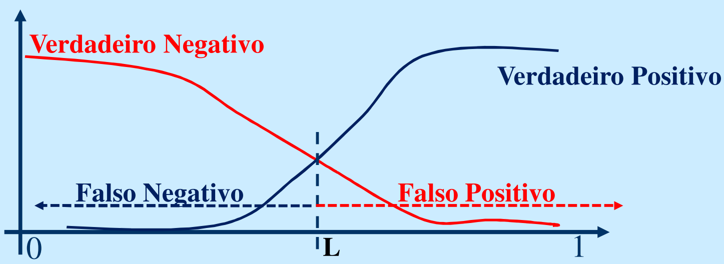
21 *Deep Learning – Conceitos e Aplicações*

Modelos de Deep Learning

Falso Positivo e Falso Negativo

Falso Positivo ➔ o modelo respondeu positivamente para um determinado exemplo cujo Target era negativo.

Falso Negativo ➔ quando o modelo responde negativamente para um determinado exemplo cujo Target era positivo.



22 *Deep Learning – Conceitos e Aplicações*

Modelos de Deep Learning

Matriz de Confusão

Matriz de Confusão

		Classes Verdadeiras	
		P	N
Classes Estimadas	P'	TP	FP
	N'	FN	TN

T ➔ True
F ➔ False

$$Acurácia = \frac{TP + TN}{P + N} \qquad precisão_P = \frac{TP}{TP + FP} \qquad precisão_N = \frac{TN}{TN + FN}$$

$$Taxa\,de\,Falso\,P = \frac{FP}{N} \qquad Taxa\,de\,Falso\,N = \frac{FN}{P}$$

23 *Deep Learning – Conceitos e Aplicações*

Modelos de Deep Learning

Acurácia em Top-K

A acurácia é calculada considerando como predição ou classificação correta se a resposta correta estiver dentre as K alternativas geradas pelo modelo.

C	E	C	E	E	C	E	E	E	C
E	C	E	E	E	E	E	E	C	E
E	E	E	E	C	E	E	E	E	E
E	E	E	E	E	E	C	E	E	E
E	E	E	E	E	E	E	C	E	E

- $A_1 = 40\%$
- $A_2 = 60\%$
- $A_3 = 70\%$
- $A_4 = 80\%$
- $A_5 = 90\%$

24 *Deep Learning – Conceitos e Aplicações*

Modelos de Deep Learning

Curva ROC

A curva ROC surgiu durante a 2ª Guerra Mundial como ferramenta usada na avaliação do desempenho dos operadores de radar – chamados de ***receiver operators***.

O objetivo era quantificar a habilidade do operador distinguir sinal útil de ruído. Esta habilidade ficou conhecida como ***Receiver Operator Characteristic*** – ROC.

Mais recentemente ela foi introduzida na área de Aprendizagem de Máquina — AM — e Mineração de Dados —MD, como uma ferramenta útil e poderosa para a avaliação de desempenho de modelos de classificação

25 *Deep Learning – Conceitos e Aplicações*

Modelos de Deep Learning

Curva ROC

		Classes Verdadeiras	
		P	N
Classes Estimadas	P'	TP	FP
	N'	FN	TN

		Classes Verdadeiras	
		P	N
Classes Estimadas	P’	p(P , P’)	p(N , P’)
	N’	p(P , N’)	p(N , N’)

$$P(X,Y) = P(X|Y)P(Y) = P(Y|X)P(X)$$

$$\boxed{Erro = p(P,N') + p(N,P') = \frac{FP + FN}{N}}$$

26 *Deep Learning – Conceitos e Aplicações*

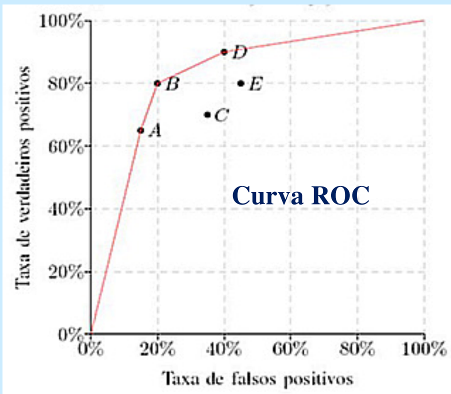
Modelos de Deep Learning

Gráfico da Curva ROC

O gráfico ROC é baseado na probabilidade de detecção, isto é, ***a taxa de verdadeiros positivos*** – *p(P,P') = TP* e na probabilidade de falsos alarmes, ou **taxa de falsos positivos** – *p(N,P') = FP*.

Para construir o gráfico ROC plota-se *FP* no eixo das ordenadas – eixo *x,* e *TP* no eixo das abscissas – eixo *y*.

O Limiar de aceitação do resultado começa em 1 (origem do gráfico) e vai sendo gradativamente diminuído até 0.



Modelos de Deep Learning

Análise da Curva ROC

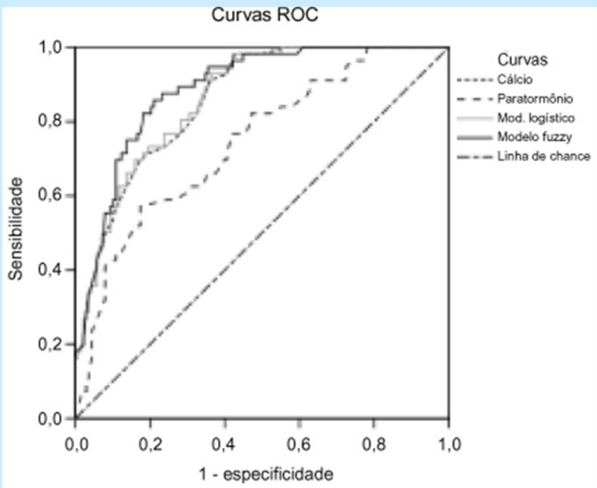
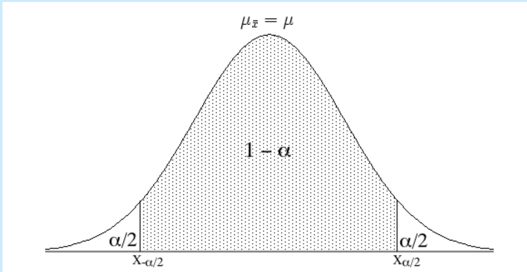


Figura 3 - Curvas ROC: Desempenho de cálcio, paratormônio, modelo logístico e modelo fuzzy na indicação de cintilografia das paratiróides.

Intervalo de Confiança

Intervalo de confiança - IC é um *indicador de precisão* da sua medida. É, também, um indicador de quão estável é a sua estimativa, isto é, quão perto a sua medição estará da estimativa original se você repetir o experimento.

Em estatística, um **intervalo de confiança** (IC) representa o intervalo estimado onde a média de um parâmetro de uma amostra tem uma certa probabilidade de ocorrer.



Intervalo de Confiança

A média amostral é um estimador não tendencioso da média populacional μ uma vez que tende a centrar-se em torno do próprio valor de μ.

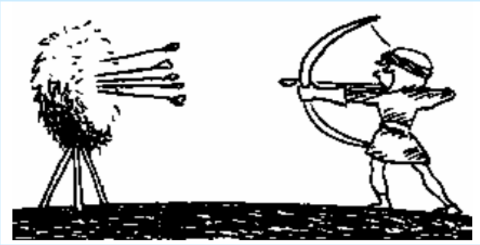
Pergunta: o que garante que um número "N" de amostras compõem uma boa estimativa da população?

Assim, associa-se a **estimativa pontual** (valor da média) a uma **estimativa intervalar** (intervalo de confiança).

Modelos de Deep Learning

Intervalo de Confiança

Suponha que numa prova de arco e flecha, uma arqueira acerta 95% das vezes em um alvo com um raio de 10cm, ou seja, erra 01 vez a cada 20 flechadas.

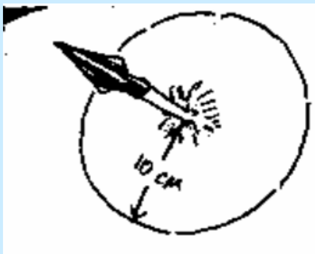


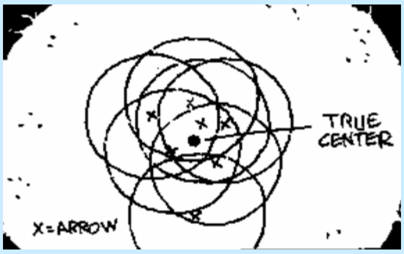
Suponha que sentado atrás do alvo tenha um árbitro, que não vê nem sabe onde está o centro do alvo, e que tentará estimar este centro com base nas flechas arremetidas pela arqueira.

Modelos de Deep Learning

Intervalo de Confiança

Sabendo do nível de habilidade da arqueira, o árbitro desenha um círculo com raio de 10 cm centrado no ponta da 1ª flecha. Ele tem 95% de confiança de que este círculo inclui o centro do alvo.





Ele se pergunta se conseguiria melhor precisão na definição do centro do alvo se desenhasse círculos semelhantes para cada uma das flechas arremetidas.

Modelos de Deep Learning

Intervalo de Confiança

Como melhorar a precisão da estimativa do centro do alvo?

Aumentando o raio do círculo ou Melhorando a mira da arqueira?

Aumentar o intervalo de confiança significa aceitar uma margem de erro maior.

33 *Deep Learning – Conceitos e Aplicações*

Modelos de Deep Learning

Intervalo de Confiança

$$\boxed{\textbf{Probabilidade}\ \{C_1 \le \mu \le C_2\} = 1 - \alpha}$$

- O intervalo (C_1, C_2) é chamado de **intervalo de confiança** da média da população.
- α é o **nível de significância (desconfiança)**.
- $100*(1-\alpha)$ é o **nível de confiança em %**.
- $1 - \alpha$ é o **coeficiente de confiança**.

34 *Deep Learning – Conceitos e Aplicações*

Cálculo do Intervalo de Confiança

A **margem de erro** pode ser calculada utilizando a seguinte fórmula:

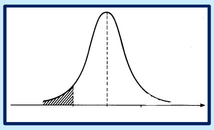
$$\boxed{ME = \pm\left(Z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right)}$$

- ✓ $Z_{\alpha/2}$ ➔ valor crítico na Distribuição Normal Padrão
- ✓ α ➔ nível de significância
- ✓ σ ➔ desvio padrão das amostras
- ✓ n ➔ número de amostras

Cálculo do Intervalo de Confiança

Exemplo:

$Pr\{x_{inf} < \mu < x_{sup}\} = 1 - \alpha = 95\%$
Onde ($\alpha = 5\%$, $\sigma = 3$ e $n = 100$)



Pr(Z < z)

- ➢ Tomando $\alpha = 0.05$
  ➔ $\alpha/2 = 0.025$
- ➢ Na tabela $\alpha/2 = 0.025$
  ➔ $Z_{\alpha/2} = -1.96$
- ➢ Na fórmula $Z_{\alpha/2} * \sigma * n^{-1/2}$
  ➔ ME = 0.588

IC ➔ $\mu \pm 0.588$

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233

Teste de Hipóteses

É um ***procedimento estatístico*** baseado na análise de uma ***Amostra*** e usado para avaliar determinados parâmetros que são desconhecidos numa ***população***.

Isto é, decidir se **determinada afirmação** sobre um ***parâmetro populacional*** é, ou não, **apoiada pela evidência** obtida a partir dos ***dados amostrais***.

Teste de Hipóteses é usado quando se tem uma informação e se quer verificar a veracidade da mesma.

https://www.youtube.com/watch?v=39dL8bk_Cxw

Teste de Hipóteses

São fundamentais os seguintes conceitos para um teste de hipótese:

- **Hipótese nula (H0)**
  - **é a hipótese assumida como verdadeira a ser testada**
- **Hipótese Alternativa (H1)**
  - **é a hipótese a ser considerada caso H0 não tenha evidência estatística**
- **Erro do Tipo I**
  - **probabilidade de se rejeitar H0 sendo ela verdadeira**
- **Erro do Tipo II**
  - **probabilidade de rejeitar H1 quando ela é verdadeira**

Modelos de Deep Learning

Teste de Hipóteses

Existem duas opções para expressar a conclusão final de um teste de hipóteses:

1ª) Comparar o valor da estatística de teste com o valor obtido a partir da distribuição teórica, específica para o teste, tomando por base um valor pré-fixado para o nível de significância.

2ª) Quantificar a chance do valor observado ou dos resultados mais extremos, sob a hipótese nula (H_0) ser verdadeira. Essa opção baseia-se na probabilidade de ocorrência de valores iguais ou superiores ao assumido pela estatística de teste. Este número é chamado de probabilidade de significância ou valor-p, frequentemente indicado apenas por p.

Valor-p e nível de significância α não são sinônimos. O valor-p é sempre obtido de uma amostra, enquanto o nível de significância é geralmente fixado antes da coleta dos dados.

Modelos de Deep Learning

Teste de Hipóteses – Nível α

- **Passo 1: Especificar as Hipóteses**

$$\begin{cases} H0 \rightarrow \mu = \mu_0 \\ H1 \rightarrow \mu < \mu_0 \end{cases} \quad \begin{cases} H0 \rightarrow \mu = \mu_0 \\ H1 \rightarrow \mu \neq \mu_0 \end{cases} \quad \begin{cases} H0 \rightarrow \mu = \mu_0 \\ H1 \rightarrow \mu > \mu_0 \end{cases}$$

- **Passo 2: Achar Z_c ou t_c Crítico**

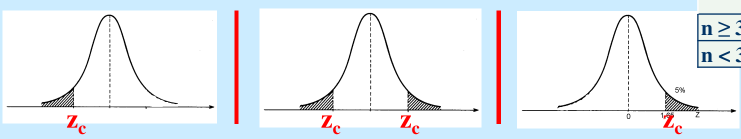


Tabela a Pesquisar

	σ	s
n ≥ 30	Z	Z
n < 30	Z	t

- **Passo 3: Achar Z ou t Estatístico**

$$Z_e = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \quad \text{ou} \quad t_e = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

- **Passo 4: Verificar se há evidência de H0 ➔ Aceitar ou Rejeitar**

Modelos de Deep Learning

Teste de Hipóteses – Nível α

Teste de Hipóteses para a Média e uma Amostra – Exemplo 1

Uma linha de produção opera com um valor médio de 16ml para o envase de um recipiente. O sobre-enchimento e o sub-enchimento são problemas sérios e a produção deve ser paralisada. De dados passados sabe-se que o desvio padrão do envase é de 0.8ml.

Um inspetor de controle de qualidade amostra 30 recipientes a cada 2 horas e, a partir deles, toma a decisão de parar a produção para calibragem dos equipamentos ou não.

Se a média amostral obtida for 15,82ml, que atitude você tomaria?

Modelos de Deep Learning

Teste de Hipóteses – Nível α

Teste de Hipóteses para a Média e uma Amostra – Exemplo 1

- **Passo 1: Especificar as Hipóteses**

$$\begin{cases} H0 \rightarrow \mu = \mu_0 = 16ml \\ H1 \rightarrow \mu \neq \mu_0 \neq 16ml \end{cases}$$

- **Passo 2: Achar Z ou t Crítico**

Bilateral com nível de significância de $\alpha = 0.05$

$\alpha = 0.025$

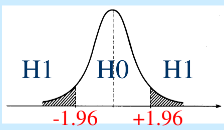


Tabela Z

	σ	s
n ≥ 30	Z	Z
n < 30	Z	t

- **Passo 3: Achar Z_e**

$$Z_e = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{15.82 - 16}{\frac{0.8}{\sqrt{30}}} = -1.23$$

Com 95% de confiança não há necessidade de paralisar a produção

- **Passo 4: Z_e está dentro dos limites críticos e, portanto, H0 não é Rejeitada**

Modelos de Deep Learning

Teste de Hipóteses – Nível α

Teste de Hipóteses para a Média e uma Amostra – Exemplo 2

Um projeto de investimento está sendo avaliado pelo método do pay-back. Uma situação envolvendo cenários futuros forneceu os seguintes tempos de Retorno do Investimento (em anos): 2.8 – 4.3 – 3.7 – 6.4 – 3.2 – 4.1 – 4.4 – 4.6 – 5.2 – 3.9.

Verifique, ao nível de significância de 5%, a hipótese de que o ROI médio seja superior a 4 anos.

Modelos de Deep Learning

Teste de Hipóteses – Nível α

Teste de Hipóteses para a Média e uma Amostra – Exemplo 2

- **Passo 1: Especificar as Hipóteses** $\begin{cases} H0 \rightarrow \mu = 4 \; anos \\ H1 \rightarrow \mu > 4 \; anos \end{cases}$

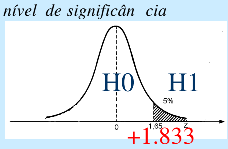
- **Passo 2: Achar Z ou t Crítico** $\begin{cases} Grau \; de \; Liberdade = n-1 \\ nível \; de \; significância \end{cases}$

Tabela t

	σ	s
n ≥ 30	Z	Z
n < 30	Z	t

Grau de liberdade = n-1

Unilateral com nível de significância de $\alpha = 0.05$



- **Passo 3: Achar t_e**

$$t_e = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{4.26 - 4}{\frac{1.02}{\sqrt{10}}} = 0.81$$

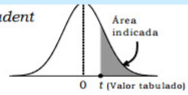
Com 95% de confiança que o ROI ocorre até os 4 anos

- **Passo 4: t_e está abaixo do limite crítico e, portanto, H0 não é Rejeitada**

Modelos de Deep Learning

Distribuição t Student

Tabela 5 Distribuição ***t*** de *Student*



gl	Área na cauda superior								
	0,25	0,10	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
1	1,000	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	0,816	1,886	2,920	4,303	6,965	9,925	14,09	22,33	31,60
3	0,765	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	0,741	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	3,833	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	3,690	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,428	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,372	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,252	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,222	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,197	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,135	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,104	3,485	3,768
24	0,685	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,078	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,689
28	0,683	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,660
30	0,683	1,310	1,697	2,042	2,457	2,750	3,030	3,385	3,646
35	0,682	1,306	1,690	2,030	2,438	2,724	2,996	3,340	3,591
40	0,681	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
45	0,680	1,301	1,679	2,014	2,412	2,690	2,952	3,281	3,520
50	0,679	1,299	1,676	2,009	2,403	2,678	2,937	3,261	3,496
z	0,674	1,282	1,645	1,960	2,326	2,576	2,807	3,090	3,291

Nota: A coluna em destaque é a mais usada.

Grau de Liberdade = 10 – 1 = 9
Nível de Significância (α) = 0.05

$\boxed{t_c = 1.833}$

Modelos de Deep Learning

Teste de Hipóteses – Nível α

Teste de Hipóteses para a Média e uma Amostra – Exemplo 3

Um fabricante de conservas anuncia que o conteúdo líquido das latas do seu produto é, em média, de 2000 gramas, com um desvio padrão de 40 gramas.

A fiscalização de pesos e medidas investigou uma amostra aleatória de 64 latas encontrando um valor médio de 1990 gramas.

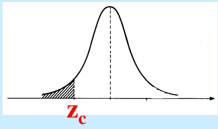
Adotando um nível de significância de 0.05, deverá o fabricante ser multado por efetuar a venda abaixo do especificado?

Modelos de Deep Learning

Teste de Hipóteses – Nível α

Teste de Hipóteses para a Média e uma Amostra – Exemplo 3

$$\begin{cases} H0: \mu = 2000 \\ H1: \mu < 2000 \end{cases}$$



$\sigma = 40$

$\bar{x} = 1990$

$n = 64$

$\alpha = 0.05$

$Z_c = -1.64$

$Z_e = -2.00$

H0 deve ser Rejeitada

O fabricante deve ser multado.

Modelos de Deep Learning

Teste de Hipóteses – p-Valor

Em um teste clássico de hipóteses, são definidas duas hipóteses, a nula (H0) e a alternativa (H1). No entanto, por utilizar para esta tomada de decisão uma amostra (uma parte da população) e não a população inteira, pode-se cometer dois tipos de erro: 1) um erro tipo I quando se rejeita H0 e H0 é verdadeira, e 2) um erro tipo II quando não se aceita H0 e H0 é falsa.

	A hipótese H_0 é verdadeira	**A hipótese H_0 é falsa**
Rejeita-se H_0	**Erro do tipo I**	**sem erro**
Não se rejeita H_0	**sem erro**	**Erro do tipo II**

A probabilidade de cometer um erro tipo I é chamada de nível de significância (α). Este nível é geralmente determinado pelo pesquisador antes da coleta dos dados, sendo tradicionalmente fixado em 0,05.

Teste de Hipóteses – p-Valor

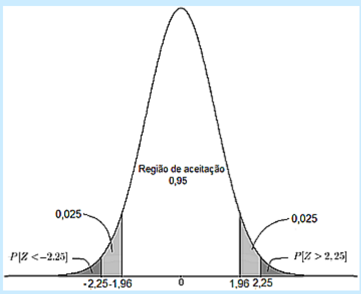
Com base nestes conceitos, define-se o **valor-p** como **a menor** escolha que poderia ser feita para o nível de significância, acima da qual ***Rejeita-se H0***.

Uma regra simplista, mas usual, define que se rejeita H0 se o valor-p for menor que α e não se rejeita H0 caso contrário.

Valor-p < α ➔ H0 é rejeitada
Valor-p > α ➔ H0 não é rejeitada

Teste de Hipóteses – p-Valor

Exemplo – Cálculo do Valor-p:



Teste Bilateral ➔ H1 $\neq \mu_0$

- ✓ $\alpha = 0.05$
- ✓ $Z_c = -1.96$
- ✓ $Z_{obs} = Z_e = -2.25$

Valor-p = $Pr\{Z < -|Z_e|\} + Pr\{Z > +|Z_e|\}$

Valor-p = 0.0244 < 0.025 ➔ H0 desse ser Rejeitada

Conclui-se que, para qualquer nível de significância maior que 0,0244, tem-se evidências para rejeitar a hipótese nula.

Teste de Hipóteses – p-Valor

P(Z<z)

z	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
-1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084

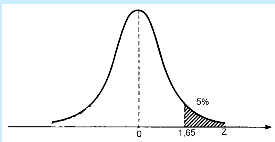
Teste de Hipóteses – p-Valor

Exercício: ***Usando o nível de significância***

Uma suinocultura usa uma ração "A" que propicia, da desmama até a idade de abate, um ganho em peso de 500 g/dia/suíno (com σ = 25 g). O fabricante de uma ração "B" afirma que nas mesmas condições, sua ração propicia um ganho de 510 g/dia (com o mesmo desvio padrão).

Se o criador tem de decidir com base em uma amostra com 50 observações, se o ganho em peso dos suínos dando a nova ração é 510 g/dia, considerando um nível de confiança de 95%, qual o teste de hipótese deve ser usado?

$$\begin{cases} H0: \mu = 500 \\ H1: \mu > 500 \end{cases} \quad \begin{array}{l} \alpha = 0.05 \Rightarrow Z_c = 1.65 \\ \sigma = 25 \therefore n = 50 \therefore \bar{x} = 510 \Rightarrow Z_e = 2.83 \end{array}$$

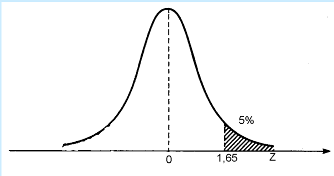


Com 95% de certeza H0 deve ser rejeitada e, portanto, há vantagem em adotar a ração B .

Modelos de Deep Learning

Teste de Hipóteses – p-Valor

Exercício: ***Usando o Valor-p***



Teste Unilateral ➔ H1 > μ_0

- ✓ $\alpha = 0.05$
- ✓ $Z_c = 1.65$
- ✓ $Z_{obs} = Z_e = 2.83$

Valor-p = $\Pr\{ Z > Z_e \} = 0.0023$

Valor-p = 0.0023 < α = 0.05 ➔ H0 deve ser Rejeitada

Há vantagem em adotar a ração B .