Classificação Linear

Lucas de Oliveira Santos
lucas.santos@lapisco.ifce.edu.br

Assuntos abordados

- ⇨ Introdução
- ⇨ Fundamentos de Classificação
- ⇨ Classificadores e Métodos
- ⇨ Análise de Componentes Principais (PCA)

Introdução

- ⇨ O que é aprendizado de máquina?
- ⇨ Reconhecimento de Padrões
- ⇨ Visão Computacional
- ⇨ Inteligência Computacional
- ⇨ Tomada de decisão

Fundamentos de Classificação

⇨ A classificação é fundamental na área do aprendizado de máquina e o seu objetivo é atribuir um ***rótulo*** ou ***classe*** a um objeto qualquer. Isso é feito baseado nas ***características*** do objeto.

⇨ As ***classes*** ou ***rótulos*** representam as distintas ***características*** às quais os objetos pertencem. Logo, o objetivo de um classificador é aprender as relações existentes entre as ***classes*** e as ***características*** de objetos.

Fundamentos de Classificação

- Existem alguns tipos de classificação, como a binária, multiclasses, dentre outras.
- Classificação binária: tem como objetivo atribuir uma entre duas classes possíveis a cada exemplo de entrada.
  - Exemplo:
    - Determinar se um tumor cerebral é benigno ou maligno;
    - Determinar se um e-mail é um spam ou não;
- Classificação multiclasses: tem como objetivo atribuir uma entre várias classes possíveis a cada exemplo de entrada.
  - Exemplo:
    - Classificar imagens de dígitos escritos à mão em sua respectivas categorias;
    - Classificar textos ou documentos em categorias de tópicos, como esportes, saúde, tecnologia, dentre outros;

Fundamentos de Classificação

a) Grupo 1

b) Grupo 2

c) Grupo 3

⇨ A qual dos grupos acima a imagem a seguir pertence ?

Fundamentos de Classificação

⇨ A sua escolha foi baseada no ***grau de semelhança*** entre a fruta desconhecida e as frutas pertencentes aos grupos 1, 2 e 3 ?

⇨ Como seu cérebro faz essa atribuição de forma instintiva ?

⇨ Certamente ele faz uma ***comparação*** entre as frutas em questão !

⇨ O objetivo dos classificadores é replicar ao máximo esse mecanismo em uma máquina.

Fundamentos de Classificação

⇨ Para realizar essas comparações, nosso cérebro precisa de alguns meios:

- ***Representação*** dos ***atributos*** ou ***características*** físicos(as) das frutas;
- ***Memorização*** das características aprendidas;
- ***Regra de decisão*** para ***classificar*** as frutas;
- Capacidade de ***aprendizado*** de novas frutas;

Fundamentos de Classificação

⇨ Vamos elencar algumas características para as frutas do nosso caso de estudo:

- ***Formato*** - { esférico, oval, alongado } = { 0, 1, 2 };
- ***Cor*** - { amarelo, vermelho, alaranjado, verde, marrom } = { 0, 1, 2, 3, 4 };
- ***Textura da casca*** - { lisa, rugosa, espinhosa } = { 0, 1, 2 };
- ***É uma fruta cítrica ?*** - { não, sim } = { 0, 1 };
- ***Possui um cheiro forte ?*** - { não, sim } = { 0, 1, 2 };

⇨ Com relação à tangerina, ela é oval, alaranjada, rugosa, é cítrica e possui cheiro forte !

Fundamentos de Classificação

- Em termos computacionais, devemos atribuir valores numéricos para esses atributos, pois os computadores só entendem números.

- Dessa forma, cada fruta (ou outro tipos de objetos) será representada como um vetor que contém cada um de suas características.
  - $\boldsymbol{x} = [\, \boldsymbol{x}_1 \; \boldsymbol{x}_2 \; \ldots \; \boldsymbol{x}_p \,]$

Fundamentos de Classificação

⇨ Podemos então representar numericamente os objetos da seguinte forma:

- ***Objeto*** Laranja: ***x*** = [***0 2 1 1 0***];
- ***Objeto*** Maçã: ***y*** = [***0 1 0 0 0***];
- ***Objeto*** Tangerina: ***z*** = [***0 2 1 1 1***];

⇨ No contexto de imagens, um ***objeto*** pode ser entendido como qualquer ***parte*** de uma imagem a qual se possa extrair pelo menos uma ***característica mensurável*** que a torne ***diferente*** do restante.

Classificadores e Métodos

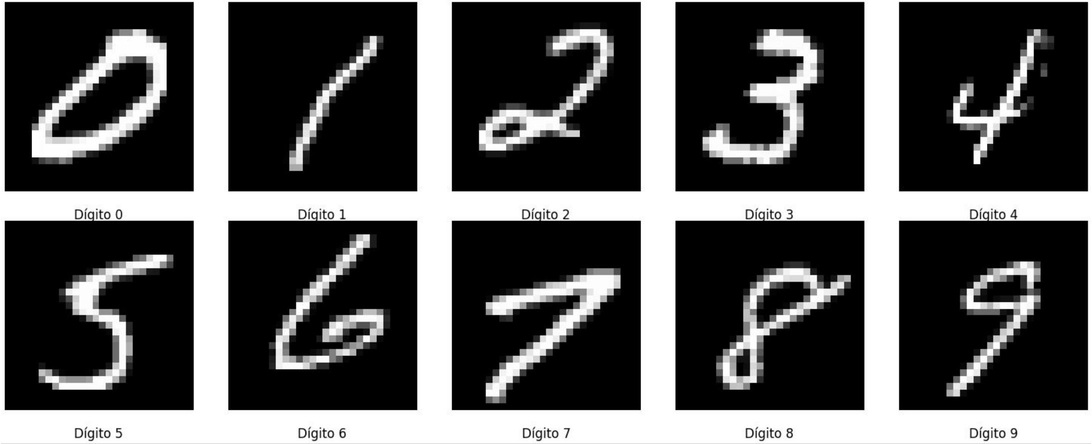
⇨ Vamos trabalhar a abordagem de classificação em imagens.

⇨ Nesse sentido, iremos resumir o problema de classificação de padrões em um modelo de regressão linear, ao qual iremos aplicar o estimador de Mínimos Quadrados - MQO.

⇨ Podemos então associar as ***classes*** como sendo um vetor de resultados ***y*** e os pixels das imagens como as características dos objetos que queremos classificar.

Classificadores e Métodos

- ⇨ Podemos estender a análise feita com as frutas para as imagens a seguir ?
- ⇨ Quais seriam as características que tornam os dígitos diferentes uns dos outros ?



Classificadores e Métodos

- Os atributos cor, tamanho e formato não podem ser efetivamente utilizados, pois os dígitos apresentam a mesma cor, um tamanho aproximado e alguns apresentam formatos parecidos.

- Dessa forma, iremos considerar os pixels de cada número como sendo os atributos deles.

Classificadores e Métodos

⇨ Iremos representar cada um desses dígitos como sendo um vetor com 10 posições (os números variam de 0 até 9). Esse método é chamado de “one-hot encoding” e é bastante usado para classificação com multi-rótulos. A imagem a seguir ilustra esse método.

$$\begin{bmatrix} C_1 \\ C_2 \\ \cdots \\ C_q \end{bmatrix} = \begin{bmatrix} [1\ 0\ 0\ \ldots\ 0]^T \\ [0\ 1\ 0\ \ldots\ 0]^T \\ \cdots \\ [0\ 0\ 0\ \ldots\ 1]^T \end{bmatrix}$$

⇨ Onde $\boldsymbol{C_1}$, $\boldsymbol{C_2}$, $\boldsymbol{C_q}$, representam as classes e os vetores respectivos representam os rótulos numéricos de cada classe.

Classificadores e Métodos

- Os coeficientes ***W*** são calculados usando-se o método MQO, que é dado pela expressão a seguir:

$$J(\beta_1, \beta_2, ..., \beta_p) = \sum_{i=1}^{n} \varepsilon_i^n = \sum_{i=1}^{2} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{j,i} \right)^2$$

- A função de soma dos quadrados é dada por: $S(\beta_0, \beta_1, ..., \beta_k) = \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{i,1} - ... - \beta_k x_{i,k})^2$

- Ela pode ser decomposta em função das derivadas parciais de cada um dos coeficientes da seguinte forma:

$$\frac{\partial S(\hat{\beta}_0, \hat{\beta}_1 ..., \hat{\beta}_k)}{\partial \beta_0} = \sum_{i=1}^{N} -2 \cdot \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - ... - \hat{\beta}_k x_{i,k} \right) = 0$$

$$\frac{\partial S(\hat{\beta}_0, \hat{\beta}_1 ..., \hat{\beta}_k)}{\partial \beta_1} = \sum_{i=1}^{N} -2 \cdot x_{i,1} \cdot \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - ... - \hat{\beta}_k x_{i,k} \right) = 0$$

$$\frac{\partial S(\hat{\beta}_0, \hat{\beta}_1 ..., \hat{\beta}_k)}{\partial \beta_k} = \sum_{i=1}^{N} -2 \cdot x_{i,k} \cdot \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i,1} - ... - \hat{\beta}_k x_{i,k} \right) = 0$$

Classificadores e Métodos

- ⇨ A formulação matricial desses conceitos é importante do ponto de vista computacional.
- ⇨ Nesse sentido, o modelo de classificação multiclasses é definido como: $Y = WX$

- ⇨ A estimativa de MQO do ponto de vista matricial é dada pela minimização da expressão: $e^T e = \sum_{i=1}^{N} e_i^2$, sendo que **e** é dado por: $e = Y - WX$. Dessa forma, podemos reescrever a equação: $e^T e = (Y - WX)^T (Y - WX)$

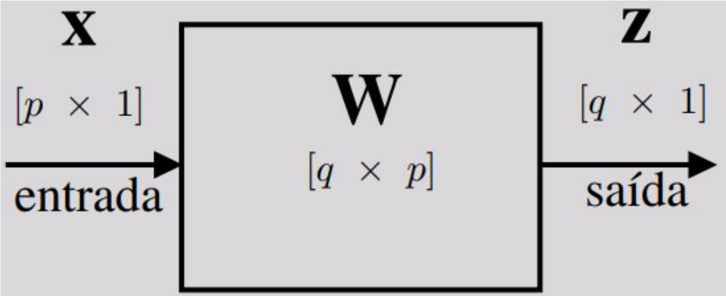
- ⇨ Por fim, a estimativa de MQO é dada por: $W = (X^T X)^{-1} (X^T Y)$. Essa equação é conhecida como o método da pseudo-inversa.

Análise das Componentes Principais - PCA

- ⇨ Em muitas aplicações reais, quando o volume de dados não é bem refinado, o modelo de classificação ou regressão pode não ser bem sucedido.
- ⇨ Dentre os vários tipos de refinamento que podem ser aplicados nos dados, iremos abordar o refinamento de dimensionalidade.
- ⇨ Ocorre quando os dados apresentam muitas características em relação ao número de amostras existentes.
- ⇨ As principais problemáticas desse tipo de comportamento nos dados são:
  - Dados muito dispersos em altas dimensões;
  - Previsões erradas;
  - Agravamento do problema de ***overfitting***;
  - Dificuldade de visualizar as relações entre as variáveis;
  - Aumento do custo computacional;

Análise das Componentes Principais - PCA

- Algumas características do PCA:
  - É um método de ***compressão de dados,*** que foi utilizado primeiramente para comprimir dados de imagens e sinais.
  - Em ***classificação de padrões***, o efeito mais vantajoso da aplicação do PCA é a ***redução nas dimensões*** dos dados.
  - O PCA aplica uma ***transformação linear*** nos dados.
- A imagem a seguir é uma representação em diagrama de blocos de uma transformação linear.



Análise das Componentes Principais - PCA

- ⇨ A equação de transformação linear do PCA é dada por: $Z = QX$
  - ***Z*** representa os dados transformados;
  - ***Q*** é a matriz de transformação;
  - ***X*** são os dados originais;
- ⇨ A equação de recuperação dos dados é dada por: $X = Q^T Z$
- ⇨ Após aplicar a transformação nos dados, as dimensões continuam as mesmas, logo, não há redução.
- ⇨ Para entender melhor como a redução de dimensão ocorre, vamos relembrar o conceito de matriz de covariância.

Análise das Componentes Principais - PCA

⇨ A ***covariância*** é uma medida estatística que nos diz como duas variáveis se ***relacionam*** entre si. Isso nos ajuda a entender se o aumento ou diminuição de uma provoca o mesmo tipo de comportamento na outra.

⇨ Se duas variáveis aumentam ou diminuem juntas, a ***covariância*** entre elas é dita positiva, caso contrário, ela é negativa.

⇨ A ***matriz de covariância*** é relevante para o problema de classificação com muitos dados porque ela nos ajuda a entender como as variáveis se relacionam umas com as outras. Ela é fundamental para utilização do PCA.

⇨ A fórmula da covariância é dada por: $S_{xy} = \frac{\sum_{i=1}^{n}(x_i \bar{x})(y_i \bar{y})}{n-1}$

Análise das Componentes Principais - PCA

⇨ Vejamos um exemplo da matriz de covariância na Tabela abaixo.

	x_i	y_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i - \bar{x})(y_i - \bar{y})$
	3	60	-5	14	-70
	4	50	-4	4	-16
	6	50	-2	4	-8
	11	40	3	-6	-18
	16	30	8	-16	-128
$\sum$	40	230	–	–	-240

⇨ Da tabela, temos que o valor da covariância ***Sxy*** = -60, logo os dados estão negativamente correlacionados. Isso indica que uma reta decrescente irá dividir os pontos.

Análise das Componentes Principais - PCA

⇨ Para entender como a redução nos dados ocorre, precisamos assumir algumas premissas:

- A matriz de ***covariância*** dos dados transformados ***Z*** é diagonal, ou seja, não existe correlação entre as componentes do vetor ***Z***;
- Os valores encontrados na diagonal da matriz ***Z*** são iguais aos valores das variâncias da matriz original ***X***.
- O cálculo das ***variâncias*** é dado por: $\sigma_x^2 = (x_i - \bar{x})^2$, $\sigma_y^2 = (y_i - \bar{y})^2$,..., $\sigma_n^2 = (n_i - \bar{n})^2$

⇨ O PCA atua sobre os dados em ***X*** para gerar um novo conjunto de dados ***Z***, cuja matriz de covariância é diagonal;

$$\mathbf{C_z} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0\cdots & 0 & \lambda_p \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0\cdots & 0 & \sigma_p^2 \end{bmatrix}$$

Análise das Componentes Principais - PCA

- Do exposto, agora podemos entender a redução de dimensionalidade que o PCA provoca nos dados com o conceito de ***Variância Total (VL)***.

$$VT = \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2$$

- A matriz diagonal obtida anteriormente possui os valores das variâncias dos dados originais, (***σ¹***, ***σ²***, …). Esses valores representam porcentagens diferentes dos dados.
- Para redução de dimensionalidade, deve-se primeiramente ordenar os valores de σ do maior para o menor. Após isso, podemos criar uma medida de quanto da informação do conjunto original está sendo representada ao escolhermos os ***q*** primeiros ***σ***.

Conclusões

- ⇨ O PCA é uma transformação linear que é aplicada aos dados de entrada;
- ⇨ As variâncias das variáveis transformadas são iguais as variâncias dos dados originais;
- ⇨ O PCA provoca uma rotação no plano original dos dados de de forma a garantir que os eixos que contenham a maior variabilidade dos dados;

Referências

- [1] W. Hines, D. Montgomery, D. Goldsman, and C. Borror, “Probabilidade e estatística na engenharia, 4a edição, ed,” LTC, Rio de Janeiro-RJ, 2006.
- [2] O. Helene, Métodos dos Mínimos Quadrados. Editora Livraria da Física, 2006.
- [3] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” Technometrics, vol. 12, no. 1, pp. 55–67, 1970.
- [4] D. Astolfi and R. Pandit, “Multivariate wind turbine power curve model based on data clustering and polynomial lasso regression,” Applied Sciences , vol. 12, no. 1, p. 72, 2021.
- [5] S. Shokrzadeh, M. J. Jozani, and E. Bibeau, “Wind turbine power curve modeling using advanced parametric and nonparametric methods,” IEEE Transactions on Sustainable Energy , vol. 5, no. 4, pp. 1262–1269, 2014.
- [6] A. E. Hoerl and R. W. Kennard, “Ridge regression: Biased estimation for nonorthogonal problems,” Technometrics , vol. 12, no. 1, pp. 55–67, 1970.

Obrigado pela atenção!

Dúvidas?

Lucas de Oliveira Santos
lucas.santos@lapisco.ifce.edu.br