线性模型（Linear Models）

代启国

大连民族大学
计算机科学与技术系

2018 年 11 月 30 日

目录

线性模型的基本形式

给定由 d 个属性描述的示例 $\mathbf{x} = (x_1; x_2; ...; x_d;)$，其中 x_i 是 $\mathbf{x}$ 在第 i 个属性上的取值。

线性模型的基本形式

给定由 d 个属性描述的示例 $\mathbf{x}=(x_1;x_2;...;x_d;)$，其中 x_i 是 $\mathbf{x}$ 在第 i 个属性上的取值。

- 线性模型（Linear model）目的是要得到这些属性的线性组合，用于预测，即

$$f(\mathbf{x})=w_1x_1+w_2x_2+...+w_dx_d+b$$

线性模型的基本形式

给定由 d 个属性描述的示例 $\mathbf{x} = (x_1; x_2; ...; x_d;)$，其中 x_i 是 $\mathbf{x}$ 在第 i 个属性上的取值。

- 线性模型（Linear model）目的是要得到这些属性的线性组合，用于预测，即
$$f(\mathbf{x}) = w_1x_1 + w_2x_2 + ... + w_dx_d + b$$
- 其一般形式为
$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$
其中，$\mathbf{w} = (w_1; w_2; ...; w_d;)$

线性模型的基本形式

给定由 d 个属性描述的示例 $\mathbf{x} = (x_1; x_2; ...; x_d;)$，其中 x_i 是 $\mathbf{x}$ 在第 i 个属性上的取值。

- 线性模型（Linear model）目的是要得到这些属性的线性组合，用于预测，即
$$f(\mathbf{x}) = w_1 x_1 + w_2 x_2 + ... + w_d x_d + b$$
- 其一般形式为
$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b$$
其中，$\mathbf{w} = (w_1; w_2; ...; w_d;)$
- 参数 $\mathbf{w}$ 和参数 b 是需要学习的参数

线性模型的特点

线性模型

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$

其中，$\mathbf{w} = (w_1; w_2; ...; w_d;)$

- 模型形式简单、易于建模
- 很多复杂的非线性（nonlinear）机器学习模型都是以线性模型为基础

线性模型的特点

线性模型

$$f(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b$$

其中，$\mathbf{w} = (w_1; w_2; ...; w_d;)$

- 模型形式简单、易于建模
- 很多复杂的非线性（nonlinear）机器学习模型都是以线性模型为基础
- 具有很好的可解释性（comprehensibility）
  - $f_{yes}(\mathbf{x}) = 0.2 \cdot x_1 + 0.5 \cdot x_2 + 0.3 \cdot x_3 + 1.35$

线性回归（Linear Regression）

给定数据集

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_m, y_m)\}$$

其中，$\mathbf{x}_i = (x_{i1}; x_{i2}; \cdots; x_{id})$，$y_i \in \mathbb{R}$

- 线性回归（Linear regression）试图学得一个线性模型以预测实数值的输出标记

$$f(\mathbf{x}_i) = \mathbf{w}\mathbf{x}_i + b$$

使得

$$f(\mathbf{x}_i) \simeq y_i$$

线性回归（Linear Regression）

给定数据集

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_m, y_m)\}$$

其中，$\mathbf{x}_i = (x_{i1}; x_{i2}; \cdots; x_{id})$，$y_i \in \mathbb{R}$

- 线性回归（Linear regression）试图学得一个线性模型以预测实数值的输出标记

$$f(\mathbf{x}_i) = \mathbf{w}\mathbf{x}_i + b$$

使得

$$f(\mathbf{x}_i) \simeq y_i$$

如何确定 $\mathbf{w}$ 和 b

线性回归（Linear Regression）

给定数据集

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_m, y_m)\}$$

其中，$\mathbf{x}_i = (x_{i1}; x_{i2}; \cdots; x_{id})$，$y_i \in \mathbb{R}$

- 线性回归（Linear regression）试图学得一个线性模型以预测实数值的输出标记

  $$f(\mathbf{x}_i) = \mathbf{w}\mathbf{x}_i + b$$

  使得

  $$f(\mathbf{x}_i) \simeq y_i$$

如何确定 $\mathbf{w}$ 和 b

- 关键：如何衡量 $f(\mathbf{x})$ 和 y 之间的差别！

给定数据集

$$D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \cdots, (\mathbf{x}_m, y_m)\}$$

其中，$\mathbf{x}_i = (x_{i1}; x_{i2}; \cdots; x_{id})$，$y_i \in \mathbb{R}$

- 线性回归（Linear regression）试图学得一个线性模型以预测实数值的输出标记

$$f(\mathbf{x}_i) = \mathbf{w}\mathbf{x}_i + b$$

使得

$$f(\mathbf{x}_i) \simeq y_i$$

如何确定 $\mathbf{w}$ 和 b

- 关键：如何衡量 $f(\mathbf{x})$ 和 y 之间的差别！
- 均方误差是回归任务中常用的性能度量指标，因此让均方误差最小化

线性回归（Linear Regression）

为简便，我们先以样本中只有一个属性为例，即一元线性回归

$$
\begin{aligned}
(w^*, b^*) &= \arg\min_{(w,b)} \sum_{i=1}^{m} (f(x_i) - y_i)^2 \\
&= \arg\min_{(w,b)} \sum_{i=1}^{m} (y_i - wx_i - b)^2
\end{aligned}
$$

线性回归（Linear Regression）

为简便，我们先以样本中只有一个属性为例，即一元线性回归

$$
\begin{aligned}
(w^*, b^*) &= \arg\min_{(w,b)} \sum_{i=1}^{m} (f(x_i) - y_i)^2 \\
&= \arg\min_{(w,b)} \sum_{i=1}^{m} (y_i - wx_i - b)^2
\end{aligned}
$$

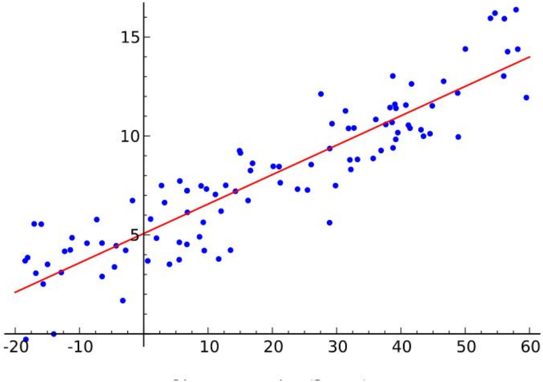
"最小二乘法"（Least square method）

- 通过均方误差最小化来进行模型求解的方法

线性回归（Linear Regression）

“最小二乘法”（Least square method）

- 通过均方误差最小化来进行模型求解的方法
- 本质上，寻找一条直线，使所有样本到直线上的欧式距离之和最小



参数估计（parameter estimation）

- 求解 w 和 b，使得 $E_{(w,b)} = \sum_{i=1}^{m} (y_i - wx_i - b)^2$ 最小化的过程

参数估计（parameter estimation）

- 求解 w 和 b，使得 $E_{(w,b)} = \sum_{i=1}^{m} (y_i - wx_i - b)^2$ 最小化的过程
- 将 $E_{(w,b)}$ 分别对 w 和 b 求导，并令导数等于 0，可得到 w 和 b 最优解闭式解 (解析解)

$$w = \frac{\sum_{i=1}^{m} y_i(x_i - \bar{x})}{\sum_{i=1}^{m} x_i^2 - \frac{1}{m}\left(\sum_{i=1}^{m} x_i\right)}$$

$$b = \frac{1}{m}\sum_{i=1}^{m}(y_i - wx_i)$$

其中，$\bar{x} = \frac{1}{m}\sum_{i=1}^{m} x_i$ 为 x 的均值

线性回归（Linear Regression）

一般地，样本由 d 个属性描述，即

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{11} & \cdots & x_{1d} \\ x_{21} & x_{21} & \cdots & x_{2d} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m1} & \cdots & x_{md} \end{pmatrix}$$

此时，我们需要学得

$$f(x_i) = \mathbf{w}^T \mathbf{x}_i + b$$

称为“多元线性回归”

对数几率回归（Logistic Regression）

线性回归预测值 $z = \mathbf{w}^T\mathbf{x} + b$ 是实值，需将其转换为 $0/1$ 值（二类分类）

- 单位阶跃函数

$$y = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

- 但该函数不连续、不可微，需找一个近似该函数的连续可微函数

对数几率回归（Logistic Regression）

线性回归预测值 $z = \mathbf{w}^T\mathbf{x} + b$ 是实值，需将其转换为 $0/1$ 值（二类分类）

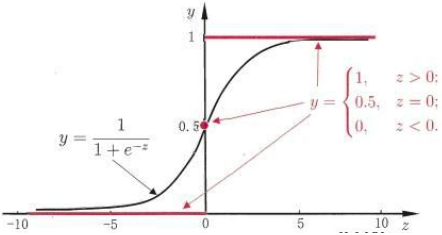
- 单位阶跃函数

$$y = \begin{cases} 1, & z \geq 0 \\ 0, & z < 0 \end{cases}$$

- 但该函数不连续、不可微，需找一个近似该函数的连续可微函数
- 对数几率函数（Logistics function）,Sigmoid 函数

$$y = \frac{1}{1 + e^{-z}}$$

$$y = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x} + b)}}$$



对数几率回归（Logistic Regression）

- 对数几率函数（Logistics function）,Sigmoid 函数

$$y = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x}+b)}}$$

对数几率回归（Logistic Regression）

- 对数几率函数（Logistics function）,Sigmoid 函数

$$y = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x}+b)}}$$

- 上式可导出

$$\ln \frac{y}{1-y} = \mathbf{w}^T\mathbf{x} + b$$

对数几率回归（Logistic Regression）

- 对数几率函数（Logistics function）,Sigmoid 函数

$$y = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x}+b)}}$$

- 上式可导出

$$\ln \frac{y}{1-y} = \mathbf{w}^T\mathbf{x} + b$$

- 若将 y 视为 $\mathbf{x}$ 为正例的可能性，则 $1 - y$ 是其为反例的可能性，则 $y/(1-y)$ 为其正负例可能性的几率

对数几率回归（Logistic Regression）

- 对数几率函数（Logistics function）,Sigmoid 函数

$$y = \frac{1}{1 + e^{-(\mathbf{w}^T\mathbf{x}+b)}}$$

- 上式可导出

$$\ln \frac{y}{1-y} = \mathbf{w}^T\mathbf{x} + b$$

- 若将 y 视为 $\mathbf{x}$ 为正例的可能性，则 $1-y$ 是其为反例的可能性，则 $y/(1-y)$ 为其正负例可能性的几率
- 对“几率”取对数，得对数几率（log odds，logit）

$$\ln \frac{y}{1-y}$$

对数几率回归（Logistic Regression）

$$\ln \frac{y}{1-y} = \mathbf{w}^T\mathbf{x} + b$$

- 用线性回归模型的预测结果逼近真实标记的对数几率，所以称为“对数几率回归”
- 虽名为回归，但实则是一种分类学习方法
- 不仅可预测”类别“，还可以得到属于类别的概率，具有很好的实用性

对数几率回归（Logistic Regression）

$$\ln \frac{y}{1-y} = \mathbf{w}^T\mathbf{x} + b$$

- 给定数据集 D，如何确定 $\mathbf{w}$和 b??

对数几率回归（Logistic Regression）

$$\ln \frac{y}{1-y} = \mathbf{w}^T\mathbf{x} + b$$

- 给定数据集 D，如何确定 $\mathbf{w}$和 b??
- 令 $y = p(y=1|\mathbf{x})$, 则有

$$p(y=1|\mathbf{x}) = \frac{e^{\mathbf{w}^T\mathbf{x}+b}}{1+e^{\mathbf{w}^T\mathbf{x}+b}}$$

$$p(y=0|\mathbf{x}) = \frac{1}{1+e^{\mathbf{w}^T\mathbf{x}+b}}$$

对数几率回归（Logistic Regression）

$$\ln \frac{y}{1-y} = \mathbf{w}^T\mathbf{x} + b$$

- 给定数据集 D，如何确定 $\mathbf{w}$和 b??
- 令 $y = p(y=1|\mathbf{x})$, 则有

$$p(y=1|\mathbf{x}) = \frac{e^{\mathbf{w}^T\mathbf{x}+b}}{1+e^{\mathbf{w}^T\mathbf{x}+b}}$$

$$p(y=0|\mathbf{x}) = \frac{1}{1+e^{\mathbf{w}^T\mathbf{x}+b}}$$

- 可以通过”极大似然法“(maximum likelihood method）对参数 $\mathbf{w}$ 和 b 进行估计

对数几率回归（Logistic Regression）

$$\ln \frac{y}{1-y} = \mathbf{w}^T\mathbf{x} + b$$

- 给定数据集 D，如何确定 $\mathbf{w}$和 b??
- 令 $y = p(y=1|\mathbf{x})$, 则有

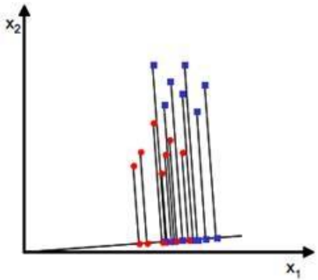
$$p(y=1|\mathbf{x}) = \frac{e^{\mathbf{w}^T\mathbf{x}+b}}{1+e^{\mathbf{w}^T\mathbf{x}+b}}$$

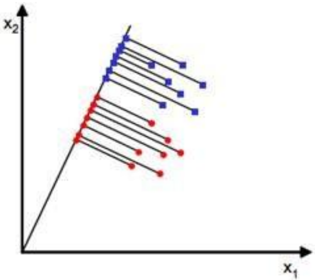
$$p(y=0|\mathbf{x}) = \frac{1}{1+e^{\mathbf{w}^T\mathbf{x}+b}}$$

- 可以通过”极大似然法“(maximum likelihood method）对参数 $\mathbf{w}$ 和 b 进行估计
- 由于实际求解过程中难以求得解析解，一般会采用梯度下降法、牛顿法等求得最优解

LDA

- 最早由 Fisher 提出，故也称 Fisher 判别分析
- 基本思想：给定训练集，设法将样例投影到一条直线上，使得同样样例的投影点尽可能近、异类样本的投影点尽可能远；
- 在对新样本进行分类时，将其投影到该直线上，根据投影点位置来判别所属类别





离散值处理

- 现实中，很多数据样本的属性值不是连续的, 而是离散的
  - 如，性别、爱好、选修课程等等
- 在很多机器学习方法（如线性模型）中，需要对离散属性值进行一定的处理，以方面模型训练与运用

,

离散值处理

- 现实中，很多数据样本的属性值不是连续的，而是离散的
  - 如，性别、爱好、选修课程等等
- 在很多机器学习方法（如线性模型）中，需要对离散属性值进行一定的处理，以方面模型训练与运用
- 对于离散值，考虑“有序”和“无序”两种情况

,

离散值处理

- 现实中，很多数据样本的属性值不是连续的，而是离散的
  - 如，性别、爱好、选修课程等等
- 在很多机器学习方法（如线性模型）中，需要对离散属性值进行一定的处理，以方面模型训练与运用
- 对于离散值，考虑“有序”和“无序”两种情况
- 有序：将其转化为连续值
  - 例如：考查课成绩的取值“优、良、中、及格、不及格”
  - 可转化为：$\{9, 8, 7, 6, 0\}$

,

离散值处理

- 现实中，很多数据样本的属性值不是连续的，而是离散的
  - 如，性别、爱好、选修课程等等
- 在很多机器学习方法（如线性模型）中，需要对离散属性值进行一定的处理，以方面模型训练与运用
- 对于离散值，考虑“有序”和“无序”两种情况
- 有序：将其转化为连续值
  - 例如：考查课成绩的取值“优、良、中、及格、不及格”
  - 可转化为：$\{9, 8, 7, 6, 0\}$
- 无序：转化为 k 维向量
  - 交通工具：火车、飞机、轮船
  - 构建 3 维向量：(1,0,0),(0,1,0),(0,0,1)
  - 对于无序离散属性，将其直接转化为连续值，会对模型训练产生误导

,

实验 1：线性回归

- 下载安装 Anaconda3
- 准备数据（可到 UCI 上下载你喜欢的数据）
- 数据预处理
- 声明学习器，并进行训练
- 模型验证
- 输出结果
- 参考：http://www.cnblogs.com/pinard/p/6016029.html