智能显示场景下的机器学习入门

黄帅

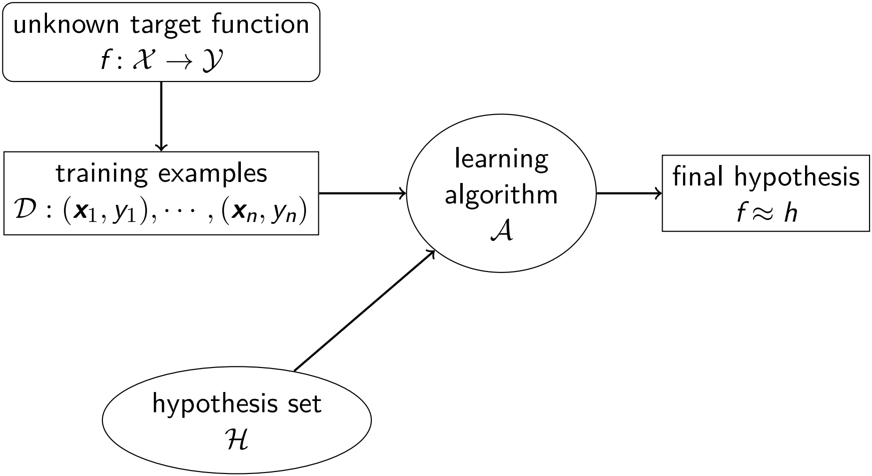
2018 年 1 月 9 日

定义

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. (by *Mitchell*)

对于某类任务 T 和性能度量 P，如果一个计算机程序在 T 上以 P 衡量的性能随着经验 E 而自我完善，那么我们称这个计算机程序在从经验 E 中学习。

- 任务 T：机器学习所关注的任务是人类很难通过总结具体规则进行解决的任务。
- 性能度量 P：为了可以量化地衡量机器学习算法的表现，通常根据不同的任务类型设计相应的性能度量方式。
- 经验 E：根据学习过程中获取到的经验种类，机器学习可以分为有监督学习和无监督学习两大类。


unknown target function
$f : \mathcal{X} \to \mathcal{Y}$
training examples
$\mathcal{D} : (\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_n, y_n)$
learning
algorithm
$\mathcal{A}$
final hypothesis
$f \approx h$
hypothesis set
$\mathcal{H}$

给定数据集 $D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\cdots,(\boldsymbol{x}_m,y_m)\}$，其中 $\boldsymbol{x}_i=(x_{i1};x_{i2};\cdots;x_{id}),y_i\in\mathbb{R}$。
“线性回归”(linear regression) 试图学得

模型定义

$$f(\boldsymbol{x})=\boldsymbol{w}^T\boldsymbol{x}+b \text{使得} f(\boldsymbol{x}_i)\approx y_i$$

其中，$\boldsymbol{x}$ 是样本中提取出的特征向量；$\boldsymbol{w}$ 和 b 是待学习参数。

- 任务 T: 回归问题，预测连续值 y_i
- 性能度量 P: 均方误差 $\frac{1}{m}\sum_{i=1}^{m}(y_i-\boldsymbol{w}\boldsymbol{x}_i)^T(y_i-\boldsymbol{w}\boldsymbol{x}_i)$
- 经验 E: 权重参数 w_i

优化过程为：令

$$\hat{\boldsymbol{w}} = (\boldsymbol{w}; b)$$

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} & 1 \\ x_{21} & x_{22} & \cdots & x_{2d} & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{md} & 1 \end{bmatrix}$$

对应的标记也写成向量形式

$$\boldsymbol{y} = (y_1; y_2; \cdots; y_m)$$

则优化目标为

$$\hat{\boldsymbol{w}}^* = \arg\min_{\hat{\boldsymbol{w}}} (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{w}})^T (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{w}})$$

令 $E_{\hat{\boldsymbol{w}}} = (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{w}})^T (\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{w}})$，对 $\hat{\boldsymbol{w}}$ 求导得：

$$\frac{\partial E_{\hat{\boldsymbol{w}}}}{\partial \hat{\boldsymbol{w}}} = 2\boldsymbol{X}^T (\boldsymbol{X}\hat{\boldsymbol{w}} - \boldsymbol{y})$$

如果 $\boldsymbol{X}^T\boldsymbol{X}$ 为正定矩阵，那么有

$$\hat{\boldsymbol{w}}^* = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

对应的线性回归模型为

$$f(\hat{\boldsymbol{x}}_i) = \hat{\boldsymbol{x}}_i^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

Q: 如果 $\boldsymbol{X}^T\boldsymbol{X}$ 不是正定矩阵？会有多个解，此时由模型的**归纳偏好**决定采用哪个解，常见的做法是引入**正则化项**。

如果要做的是分类任务应该怎么办？只需要找一个单调可微函数将分类任务的真实标记 y 与线性回归模型的预测值联系起来.

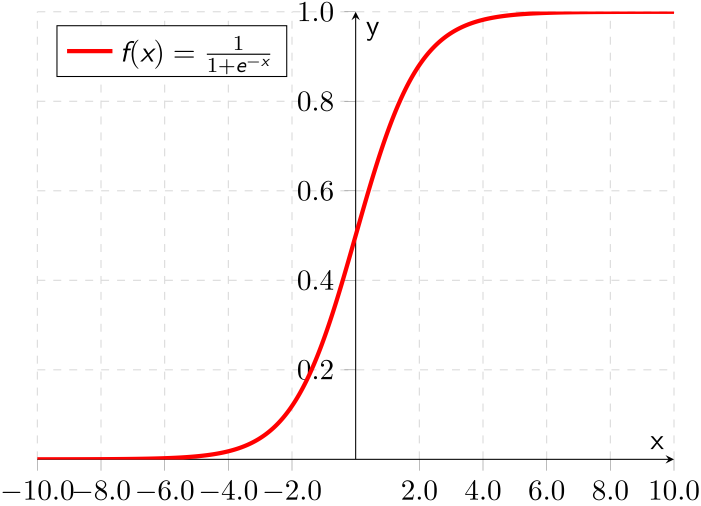
对数几率回归模型

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

其中，$\boldsymbol{x}$ 是样本中提取出的特征向量；$\boldsymbol{w}$ 和 b 是待学习参数。

对数几率回归又称 *Logistic Regression*, 虽然名字中带着回归两个字，但它是一个分类模型。

- 任务 T: 分类任务
- 性能度量 P: 准确率 (精准率, 召回率)
- 经验 E: 权重参数 w_i


$f(x) = \frac{1}{1+e^{-x}}$
y
x
1.0
0.8
0.6
0.4
0.2
−10.0
−8.0
−6.0
−4.0
−2.0
2.0
4.0
6.0
8.0
10.0

由定义可得,

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

其中, $\frac{y}{1-y}$ 被称为**几率**, 反映了样本 $\boldsymbol{x}$ 作为正例的相对可能性. 对几率取对数则得到**对数几率**.

如何求取参数 $\boldsymbol{w}, b$? 设

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

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

通过极大化对数似然

$$\mathcal{L}(\boldsymbol{w}, b) = \sum_{i=1}^{m} \ln p(y_i|\boldsymbol{x}_i; \boldsymbol{w}, b)$$

求取相应的参数 $(\boldsymbol{w}, b)$.

真实情况	预测结果	
	正例	反例
正例	TP(真正例)	FN(假反例)
反例	FP(假正例)	TN(真反例)

- 精准率 (查准率)

$$P = \frac{TP}{TP + FP}$$

- 召回率 (查全率)

$$R = \frac{TP}{TP + FN}$$

GitHub Repository