HW1

asandstar@github

2022 年 10 月 26 日

1 简述机器学习的类别，监督学习过程中涉及哪些要素

(1) 机器学习可以分为监督学习、无监督学习、强化学习、半监督学习与主动学习。
监督学习指从标注数据中学习预测模型的机器学习问题。
无监督学习指从无标注数据中学习预测模型的机器学习问题。
强化学习指智能系统在与环境的连续互动中学习最优行为策略的机器学习问题。
半监督学习指利用标注数据和未标注数据学习预测模型的机器学习问题。
主动学习指机器不断给出实例让教师进行标注，再利用标注数据学习预测模型的机器学习问题。
(2) 监督学习过程中涉及的要素为数据、模型、策略和算法。其中三要素为：模型、策略和算法。

2 现有以下数据点 (x_i,y_i) ，试用二次函数拟合，在最大似然准则下估计系数，并计算拟合残差 (0,-0.13), (0.2,0.09), (0.4,0.59), (0.6,1.62), (0.8,2.5), (1,3.4)

已知 $X = [0\ 0.2\ 0.4\ 0.6\ 0.8\ 1]$，
$t = [-0.13\ 0.09\ 0.59\ 1.62\ 2.5\ 3.4]$
$w* = arg\,\underset{w}{min}\sum_{i=1}^{6}(f(x_i, w) - y_i)^2$
$\sum_{i=1}^{6}\frac{\partial(f(x_i, w) - y_i)^2}{\partial w} = 2\sum_{i=1}^{6}(f(x_i, w) - y_i)\frac{\partial f(x_i, w)}{\partial w} = 0$
由于 $\frac{\partial f(x_i, w)}{\partial w} = [x_i^2, x_i, 1]$

$$\begin{cases} \sum_{i=1}^{6}(y_i - w_2x_i^2 - w_1x_i - w_0)x_i^2 = 0 \\ \sum_{i=1}^{6}(y_i - w_2x_i^2 - w_1x_i - w_0)x_i = 0 \\ \sum_{i=1}^{6}(y_i - w_2x_i^2 - w_1x_i - w_0) = 0 \end{cases}$$

$$\begin{bmatrix} \sum_{i=1}^{6} y_ix_i^2 \\ \sum_{i=1}^{6} y_ix_i \\ \sum_{i=1}^{6} y_i \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{6} x_i^4 & \sum_{i=1}^{6} x_i^3 & \sum_{i=1}^{6} x_i^2 \\ \sum_{i=1}^{6} x_i^3 & \sum_{i=1}^{6} x_i^2 & \sum_{i=1}^{6} x_i \\ \sum_{i=1}^{6} x_i^2 & \sum_{i=1}^{6} x_i & 6 \end{bmatrix} \cdot \begin{bmatrix} w_2 \\ w_1 \\ w_0 \end{bmatrix}$$

```python
import numpy as np
from numpy import *
res1 = res2 = 0
m1 = m2 = m3 = 0
count1 = count2 = count3 =0
count4 = count5 = 0
x = [0, 0.2, 0.4, 0.6, 0.8, 1]
y = [-0.13, 0.09, 0.59, 1.62, 2.5, 3.4]
for i in range(6):
    res1 = y[i] * x[i] * x[i]
    res2 = y[i] * x[i]
    m1 = x[i] * x[i] * x[i] * x[i]
    m2 = x[i] * x[i] * x[i]
    m3 = x[i] * x[i]
    count1 = count1 + res1
    count2 = count2 + m1
    count3 = count3 + m2
    count4 = count4 + res2
    count5 = count5 + m3
res3 = sum(y)
m4 = sum(x)
m5 = 6
yx = np.array([[count1], [count4], [res3]])
xx = np.array([[count2, count3, count5],
               [count3, count5, m4],
               [count5, m4, m5]])
xx_inv = np.linalg.inv(xx)
w2 = w1 = w0 = 0
w = [w2, w1, w0]
w = np.dot(xx_inv, yx)
print(w)
y_see = np.array(y)
s = 0
y_p = {}
for i in range(6):
    y_p[i] = w[0]*x[i]*x[i]+w[1]*x[i]+w[2]
    s = s+y_see[i]-y_p[i]
print(s)
```

解得 $w_2 = 2.19642857,\ w_1 = 1.505,\ w_0 = -0.21285714$

$y(x, \boldsymbol{w}) = -0.21285714 + 1.505x + 2.19642857x^2$

拟合残差 $\varepsilon = \sum_{i=1}^{6}(t - y(x_i, \boldsymbol{w}))$

计算得 $\varepsilon = 3.06421555\times10^{-14}$

3 题 2 中设 y_i 的观测误差服从高斯分布 N(0,0.1)，二次函数的各系数均服从先验分布 N(0,0.3)，试在最大后验准则下重新估计各系数

$y_i = f(w, x) + \delta$

$\delta \sim N(0, \sigma_1),\ \sigma_1 = 0.1$

$w \sim N(0, \sigma_2),\ \sigma_2 = 0.3$

后验概率 $P(w) = exp(-\dfrac{w^2}{2\sigma_2^2})\prod\limits_{i=1}^{6} exp(-\dfrac{(y_i - f(w, x_i))^2}{2\sigma_1^2})$

两边取对数后对 w_j 求导得 $\dfrac{\mathrm{d}ln\ P}{\mathrm{d}w_j} = \dfrac{\mathrm{d}[(-\frac{\boldsymbol{w}^2}{2\sigma_2^2}) + \sum_{i=1}^{6}(-\frac{(y_i - f(\boldsymbol{w}, x_i))^2}{2\sigma_1^2})]}{\mathrm{d}w_j}$

$= \dfrac{\mathrm{d}[(-\frac{\boldsymbol{w}^2}{2\sigma_2^2}) + \sum_{i=1}^{6}(-\frac{(y_i - w_2 x_i^2 - w_1 x_i - w_0)^2}{2\sigma_1^2})]}{\mathrm{d}w_j}$

令上式 $= 0$ 解得

$\sum_{i=1}^{6} x_i^2(y_i - w_2 x_i^2 - w_1 x_i - w_0) - \frac{\sigma_1^2}{\sigma_2^2} w_2 = 0$

$\sum_{i=1}^{6} x_i(y_i - w_2 x_i^2 - w_1 x_i - w_0) - \frac{\sigma_1^2}{\sigma_2^2} w_1 = 0,$

$\sum_{i=1}^{6}(y_i - w_2 x_i^2 - w_1 x_i - w_0) - \frac{\sigma_1^2}{6\sigma_2^2} w_0 = 0$

$$\begin{bmatrix} \sum\limits_{i=1}^{6} y_i x_i^2 \\ \sum\limits_{i=1}^{6} y_i x_i \\ \sum\limits_{i=1}^{6} y_i \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{6} x_i^4 + \frac{\sigma_1^2}{\sigma_2^2} & \sum\limits_{i=1}^{6} x_i^3 & \sum\limits_{i=1}^{6} x_i^2 \\ \sum\limits_{i=1}^{6} x_i^3 & \sum\limits_{i=1}^{6} x_i^2 + \frac{\sigma_1^2}{\sigma_2^2} & \sum\limits_{i=1}^{6} x_i \\ \sum\limits_{i=1}^{6} x_i^2 & \sum\limits_{i=1}^{6} x_i & 6 + \frac{\sigma_1^2}{6\sigma_2^2} \end{bmatrix} \cdot \begin{bmatrix} w_2 \\ w_1 \\ w_0 \end{bmatrix}$$

上述代码需要修改的部分为 24 至 26 行

```
xx = np.array([[count2+1/9, count3, count5],
               [count3, count5+1/9, m4],
               [count5, m4, m5+1/(6*9)]])
```

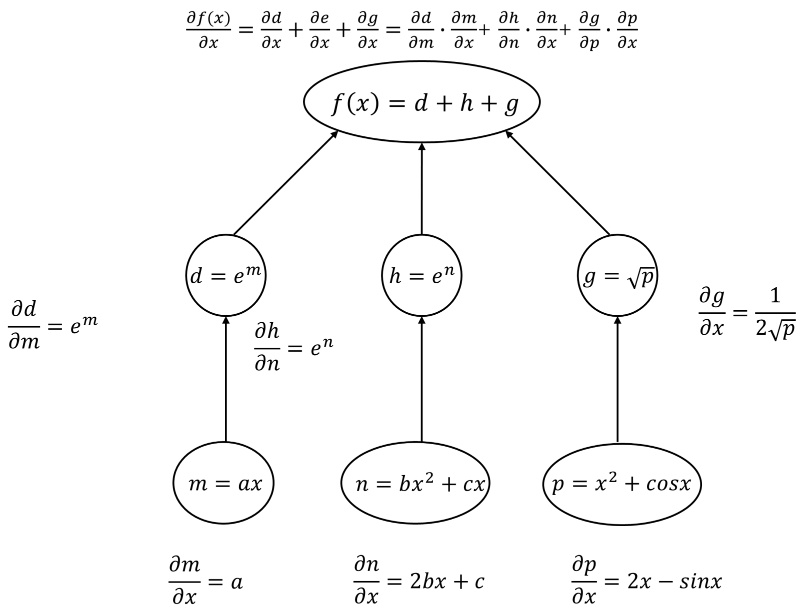
带入数值得 $w_2 = 1.82312499,\ w_1 = 1.6194856,\ w_0 = -0.13281205$

$y(x, \boldsymbol{w}) = -0.13281205 + 1.6194856x + 1.82312499x^2$

拟合残差 $\varepsilon = \sum\limits_{i=1}^{6}(t - y(x_i, \boldsymbol{w}))$

计算得 $\varepsilon = -0.00245948$

4 试画出下式的计算图，并在该图上用链式法则计算梯度 $\frac{\partial f}{\partial x}$：$f(x)=e^{ax}+e^{bx^2+cx}+\sqrt{x^2+cosx}$



$f(x)$	e^{ax}	e^{bx^2+cx}	$\sqrt{x^2+cosx}$
$\frac{\partial f(x)}{\partial x}$	ae^{ax}	$(2bx+c)e^{bx^2+cx}$	$\frac{2x-sinx}{2\sqrt{x^2+cosx}}$

$\frac{\partial f}{\partial x}=ae^{ax}+(2bx+c)e^{bx^2+cx}+\frac{2x-sinx}{2\sqrt{x^2+cosx}}$

5 试证明以下矩阵代数的等式

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{a}^\top$$

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{X}\boldsymbol{b}}{\partial \boldsymbol{X}}=\boldsymbol{a}\boldsymbol{b}^\top$$

$$\frac{\partial \boldsymbol{x}^\top \boldsymbol{B}\boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{x}^\top(\boldsymbol{B}+\boldsymbol{B}^\top)$$

证明：

(1)

$\boldsymbol{a}^\top=[a_1\ a_2\ a_3\ \cdots\ a_n],\ x=[x_1\ x_2\ x_3\ \cdots\ x_n]^\top$

因为 $\frac{\partial \boldsymbol{a}^\top \boldsymbol{x}}{\partial x_i}=\frac{\partial(\sum\limits_{j=1}^{n}a_jx_j)}{\partial x_i}=a_i$

所以 $\frac{\partial \boldsymbol{a}^\top \boldsymbol{x}}{\partial \boldsymbol{x}}=[a_1\ a_2\ a_3\ \cdots\ a_n]=\boldsymbol{a}^\top$

证毕

(2)

$\boldsymbol{a}^\top = [a_1\ a_2\ a_3\ \cdots\ a_m]$, $\boldsymbol{b} = [b_1\ b_2\ b_3\ \cdots\ b_n]^\top$

因为 $\dfrac{\partial \boldsymbol{a}^\top \boldsymbol{X} \boldsymbol{b}}{\partial X_{ij}} = \dfrac{\partial(\sum\limits_{t=1}^{m}\sum\limits_{k=1}^{n} a_t X_{tk} b_k)}{\partial X_{ij}} = a_i b_j$

所以

$$\frac{\partial \boldsymbol{a}^\top \boldsymbol{X} \boldsymbol{b}}{\partial\ \boldsymbol{X}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \ldots & b_n \end{bmatrix} = \boldsymbol{a}\boldsymbol{b}^\top$$

证毕

(3)

$B_{n\times n} = (b_{ij})_{i=1,j=i}^{n,n}$, $x = [x_1\ x_2\ x_3\ \cdots\ x_n]^\top$

因为 $\dfrac{\partial \boldsymbol{x}^\top \boldsymbol{B} \boldsymbol{x}}{\partial x_{ij}} = \dfrac{\partial(\sum\limits_{t=1}^{n}\sum\limits_{k=1}^{n} b_{tk} x_t x_k)}{\partial x_t} = \sum\limits_{k=1}^{n} b_{tk} x_k$

所以

$$\frac{\partial \boldsymbol{x}^\top\ \boldsymbol{B}\boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{bmatrix} b_{11}x_1 + b_{12}x_2 + & \ldots & +b_{1n}x_n \\ b_{21}x_1 + b_{22}x_2 + & \ldots & +b_{2n}x_n \\ \vdots & & \\ b_{n1}x_1 + b_{n2}x_2 + & \ldots & +b_{nn}x_n \end{bmatrix} + \begin{bmatrix} b_{11}x_1 + b_{21}x_2 + & \ldots & +b_{n1}x_n \\ b_{12}x_1 + b_{22}x_2 + & \ldots & +b_{n2}x_n \\ \vdots & & \\ b_{1n}x_1 + b_{2n}x_2 + & \ldots & +b_{nn}x_n \end{bmatrix}$$

$$= \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^\top \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \ldots & b_{nn} \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}^\top \begin{bmatrix} b_{11} & b_{21} & \ldots & b_{n1} \\ b_{12} & b_{22} & \ldots & b_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1n} & b_{2n} & \ldots & b_{nn} \end{bmatrix} = \boldsymbol{x}^\top(\boldsymbol{B} + \boldsymbol{B}^\top)$$

证毕