hw1

guangbo yu ID: 7262891072

September 21, 2016

homework1

1 Density Estimation

1.1

1.1.1

when $\beta = 1$ the β function is

$$f(x) = \frac{1}{Beta(\theta, 1)}\ x^{\theta-1} = \theta\ x^{a-1}$$

$$L(\theta) = \theta^n \prod_{i=1}^{n} x_i^{\theta-1}$$

$$\log^{L(\theta;x)} = n\log^{\theta} + (\theta-1)\sum_{i=1}^{n} \log_i^x$$

$$\frac{\partial \log^L(\theta)}{\partial\theta} = \frac{n}{a} + \sum_{i=1}^{n} \log_i^x = 0$$

$$\theta = \frac{-n}{\sum_{i=1}^{n} \log_i^x}$$

1.1.2

$$f(x) = (\frac{1}{\sqrt{2\pi\theta}})^n e^{-\frac{(x-\theta)^2}{2\theta}}$$

$$\log^L = -\frac{n}{2}\log^{2\pi\theta} - \sum_{i=1}^{n} \frac{x_i^2}{2\theta} - \frac{n\theta}{2} + \sum_{i=1}^{n} x_i$$

$$\frac{\partial \log^L}{\partial\theta} = -\frac{n}{2\theta} + \sum_{i=1}^{n} \frac{x_i^2}{2\theta^2} - \frac{n}{2} = 0$$

$$\theta_1 = \frac{-n + \sqrt{n^2 - 4n\sum_{i=1}^{n} x_i^2}}{2n}$$
$$\theta_2 = \frac{-n - \sqrt{n^2 - 4n\sum_{i=1}^{n} x_i^2}}{2n}$$

1.2

1.2.1

$$E(f(\hat{x})) = \frac{1}{n}\sum_{i=1}^{n} E(\frac{1}{h}k(\frac{x_i - x}{h}))$$

let $\mu = \frac{x_i - x}{h}$ we have $dt = hd\mu$

$$E(f(\hat{x})) = \int \frac{1}{h} h f(x + h\mu)k(\mu)\, d\mu$$

$$E(f(\hat{x})) = \int f(x + h\mu)k(\mu)\, d\mu$$

$$E(f(\hat{x})) = \int \frac{1}{h}k(\frac{x - t}{h})f(t)\, dt$$

1.2.2

$$f(x + hz) = f(x) + f^{(1)}(x)hz + \frac{1}{2}f^{(2)}(x)h^2z^2 + \cdots + \frac{1}{v!}f^{(x)}h^vz^v + o(h^v)$$

1.2.3

let the j-moment of kernel is $k_j(k) = 0$ for j < v

$$E(f(\hat{x})) = \int f(x + h\mu)k(\mu)$$

$$E(f(\hat{x})) = f(x) + f^{(1)}(x)hk_1(k) + \frac{1}{2}f^{(2)}(x)h^2k_2(k) + \cdots + \frac{1}{v!}f^v(x)h^vk_v(k) + o(h^v)$$
$$E(f(\hat{x})) = f(x) + \frac{1}{v!}f^{(v)}(x)h^vk_v(k) + o(h^v)$$
$$Bias(f(\hat{x})) = E(f(\hat{x})) - f(x)$$
$$Bias(f(\hat{x})) = \frac{1}{v!}f^v(x)h^vk_v(k) + o(h^v)$$

2 Naive Bayes

2.1

$$P(Y = C_k|X) = \frac{\prod_{i=1}^{D} P(x_i|y = c_k)P(y = c_k)}{\sum_k = 0^1 P(y = y_k) \prod_{i=1}^{D}}$$

$$P(Y = 1|X) = \frac{\prod_{i=1}^{D} P(x_i|y = 1)P(y = 1)}{P(y = 1) \prod_{i=1}^{D} P(x_i|y = 1) + P(y = 0) \prod_{i=1}^{D} P(x_i|y = 0)}$$

$$P(X_i|Y = y_k) \ \frac{1}{\sqrt{2\pi\sigma}} \exp - \frac{(x - \mu_{ik})^2}{2\sigma^2}$$

$$P(Y = 1) = \pi$$

$$= \frac{\prod_{i=1}^{D} \pi \frac{1}{\sqrt{2\pi\sigma_i}} \exp - \frac{(x-\mu_{i1})^2}{2\sigma_i^2}}{\prod_{i=1}^{D} \pi \frac{1}{\sqrt{2\pi\sigma_i}} \exp - \frac{(x-\mu_{i1})^2}{2\sigma_i^2} + \prod_{i=1}^{D} (1 - \pi) \frac{1}{\sqrt{2\pi\sigma_i}} \exp - \frac{(x-\mu_i 0)^2}{2\sigma_i^2}}$$

$$= \frac{1}{1 + e^{\ln \frac{1-\pi}{\pi} + \frac{\sum_{i=1}^{D} ((x_i-\mu_{i1})^2 - (x_i-\mu_{i0})^2)}{2\sigma_1^2}}}$$

$$= \frac{1}{1 + \exp(-w_0 + w^T x)}$$

$$w_0 = \ln \frac{\pi - 1}{\pi}$$

$$w_t = \sum_i (\frac{\mu_{i0} - \mu_{i1}}{\sigma_i^2} X_i + \frac{\mu_{i1}^2 - \mu_{i0}^2}{2\sigma_i^2})$$

2.2

Likehood:

$$L(\pi_k, \mu_{jk}, \sigma_{jk} = \sum_{i=1}^{N} [\log^{P(Y=y_i)) \prod_{j=1}^{D} \frac{1}{\sqrt{2\pi\sigma_{jk}}} \exp - \frac{(x_{ij}-\mu_{ij})^2}{2\sigma_{jk}}}])$$

$$= \sum_{k=1}^{2} n_k \log \pi_k + \sum_{k=1}^{2} \sum_{i=1}^{N} I(y_i = k)[\sum_{j=1}^{D} (-\log \sqrt{2\pi\sigma_{jk}} - \frac{(x_{ij} - \mu_{jk})^2}{2\sigma_{jk}})]$$

$$\frac{\partial L}{\partial L} = n_k \frac{1}{\pi_k} + \lambda = 0$$

$$\pi_k = -\frac{n_k}{\lambda}$$

$\sum_k \pi_k = 1$, $\lambda = -N$

$$\hat{\pi_k} = \frac{n_k}{N}$$

$$\frac{\partial L}{\partial \mu_{jk}} = \sum_{i=1}^{N} I(Y_i = k)x_{ij} - n_k\mu_{jk} = 0$$

$$\hat{\mu_{jk}} = \frac{\sum_{i=1}^{N} I(Y_i = k)x_{ij}}{n_k}$$

$$\frac{\partial L}{\partial \sigma_{jk}} = \sum_{i=1}^{N} I(Y_i = k)(-\frac{1}{2\sigma_{jk}} + \frac{(x_{ij} - \mu_{jk})^2}{2\theta_{jk}^2}) = 0$$

$$\hat{\sigma_{jk}} = \frac{1}{n_k}\sum_{i=1}^{N} I(Y_i = k)(x_{ij} - \hat{\mu_{jk}})^2$$

3 KNN

3.1

$\hat{x} = \frac{1}{n}\sum x = 12.6$ $\hat{y} = \frac{1}{n}\sum y = 12.3$

$$S_x = \sqrt{\frac{1}{n-1}\sum(x_i - \hat{x})^2} = 20.7$$

$$S_y = \sqrt{\frac{1}{n-1}\sum(y_i - \hat{y})^2} = 25.9$$

according to $x = \frac{x-\hat{x}}{S_x}$, $y = \frac{y-\hat{y}}{S_y}$ we get: Math(-0.616, 1.416) (-0.96, 0.76) (-1.05, 1.34)

EE (0.78, 0) (1.75, 0.72) (1.17, 1)

CS (-0.23, -0.13) (0, -0.51) (-0.9 -0.59) (-1.63, 0)

ECON (0.68, -1.71) (0.3, -1.02) (0.68, -1.25)

Caculate L_1:

Math:

2.59

2.27

2.94

EE:

0.63

2.32

2.02

CS:

0.65

0.66

1.64

2.18
ECON:
1.84
0.87
1.38

sort the nearest 5: EE $<$ CS $<$ CS $<$ ECON $<$ ECON
$K = 1$: result is Electrical engineering $K = 5$: result is computer science
Caculate L_2
Math:
1.89
1.62
2.08

EE:
0.474
0.68
1.45

Cs:
0.58
0.468
1.31
1.99

Econ:
1.55
0.82
1.10

CS $<$ EE $<$ CS $<$ Econ $<$ Econ
K = 1, result is computer science
K = 5, result is computer science
$\frac{3}{4}$'s condition is computer scienc $\frac{1}{4}$'s condition is electical engineering

3.2

3.2.1

$P(x) = \sum_{c=1}^{N} P(X|\hat{Y} = C)P(\hat{Y} = C) = \sum_{i=1}^{N} \frac{K_c N_c}{N_c N} = \sum_{i=1}^{N} \frac{K_c}{V_N}$

$$= \frac{K}{V_N}$$

3.2.2

$P(\hat{Y} = C|X) = \frac{P(X|Y=C)P(Y=C)}{\sum_{(} i=1)^N P(X|Y=C_i)P(Y=C_i)} = \frac{\frac{K_c N_c}{N_c V N}}{\frac{K}{VN}} = \frac{K_c}{K}$

4 Decision Tree

4.1

Wether:

sunny:

$$H(Y|X_1) = -\frac{28}{100}(\frac{23}{28}\ln\frac{23}{28} - \frac{5}{28}\ln\frac{5}{28}) = 0.13$$

rainy:

$$H(Y|X_2) = -\frac{72}{100}(\frac{50}{72}\ln\frac{50}{72} - \frac{22}{72}\ln\frac{22}{72}) = 0.44$$

$$H_1(Y|X) = H(Y|X_1) + H(Y|X_2) = 0.57$$

Traffic:

$$H(Y|X_1) = -\frac{73}{100}(\frac{73}{73}\ln\frac{73}{73} - 0) = 0$$

$$H(Y|X_1) = -\frac{27}{100}(\frac{27}{27}\ln\frac{73}{73} - 0) = 0$$

$$H_2(Y|X) = 0$$

$H_1(Y|X) > H_2(Y|X)$ $Gain(1) < Gain(2)$

we should choose traffic condition firstly since its has larger gain

4.2

N_1 is more accurate since decision tree is not normal distribution

4.3

We can compare $(1 - p_k)$ and $\log^{p_k}$ for every k

let $f(p_k) = 1 - p_k - (-\log^{p_k})$

$f^{'}(p_k) = \frac{1-p_k}{p_k}$, since $0 \leq p_k \leq 1$, so $f^{'}(p_k) > 0$

therefore $f(p_k)$ is on increasing, $f(p_k)_{max} = f(1) = 0$, so $1 - p_k \leq -\log^{p_k}$

so $\sum_{k=1}^{k} p_k(1 - p_k) \leq -\sum_{k=1}^{k} p_k \log_{p_k}$

$$Giniindex \leq Cross - entropy$$

5 Programing

5.1 Data

5.1.1

we have 9 attribute

5.1.2

Yes, all are meaningful

5.1.3

6 Classes

5.1.4

Class 2 is majority, It is Normal distribution

5.2 KNN Accuracy

5.2.1 train accuracy

$$L_1$$

K=1, acc = 0.92. K=3 acc = 0.898, K=5 acc = 0.908. K=7, acc = 0.923.

$$L_2$$

K=1, acc = 0.87. K=3 acc = 0.89, K=5 acc = 0.88. K=7, acc = 0.88.

5.2.2 Test accuracy

$$L_1$$

K=1, acc = 0.83. K=3 acc = 0.78, K=5 acc = 0.78. K=7, acc = 0.67.

$$L_2$$

K=1, acc = 0.72. K=3 acc = 0.61, K=5 acc = 0.67. K=7, acc = 0.72.