

1. (20 points) Answer the following questions as follows:

1.1. Naive Bayes algorithm assumes the feature independence. TRUE or FALSE? and Why?

Answer:

It is True:

if event A, B are independent
then $P(A|B) = P(A)$... (1)
we know $P(A|B) = \frac{P(AB)}{P(B)}$... (2)
then $\stackrel{(1)}{(2)} \Rightarrow P(A)P(B) = P(AB)$... (3)

for $P(Y|X)$ where $X = \{X_1, X_2, \dots, X_n\}$
 $P(Y|X) = P(Y) \frac{P(X|Y)}{P(X)}$... (4)

for ~~$P(X|Y)$~~ , when X_1, X_2, \dots, X_n are independent, $P(X|Y) = \prod_{i=1}^n P(x_i|y)$... (5)
equation (5) is easy to compute.

This is the reason why it is called naive bayes when we applying Bayesian formula.

1.2. The functions f and g are both concave functions of a single variable. Neither function is necessarily differentiable. Is the function h defined by $h(x) = f(x) + g(x)$ necessarily concave, necessarily convex, or not necessarily either?

Answer:

$h(x)$ is the same as $f(x)$ and $g(x)$, they are all concave functions.

1.3. Find the angle between these two vectors, $u = (6, 2)$, and $v = (5, -2)$.

Answer:

$$u = (6, 2) \quad \|u\| = \sqrt{6^2 + 2^2} = \sqrt{40} \quad \vec{u} \cdot \vec{v} = 6\sqrt{5} - 2 \times 2 \\ v = (5, -2) \quad \|v\| = \sqrt{25 + 4} = \sqrt{29} \quad = 26 \\ \cos \theta = \frac{\vec{u} \cdot \vec{v}}{\|u\| \cdot \|v\|} = \frac{26}{\sqrt{40} \times \sqrt{29}} \approx 0.7634 \\ \theta \approx \arccos(0.7634) \\ \approx 40.24^\circ$$

1.4. Suppose we were interested in determining if there were differences in the average prices among two automobile companies. We randomly pick ten cars to compare at both companies. Which statistical procedure would be best to use for this study and why?

Answer:

We can apply hypothesis testing on two sets of samples to test whether they are independent or related no matter what PDF do they obey. Before start a parametric hypothesis testing, we need to know the statistic mean of the sample set. But Central Limit Theorem holds especially true for sample sizes over 30. So the sample size should be noticed because there are not enough observations on both RV, each of them contains ten observations. As a result, non-parametric hypothesis testing should be considered.

Considering two sets of car prices as two sequences and applying Mann-Whitney Test of Independence is a good hypothesis testing opinion on both equal or unequal size of sample sets.

2. (20 points)

Assume that when deciding whether an applicant is low-risk (c0), medium-risk (c1), or high-risk (c2) for bank loans, the bank only considers the amount of money in the saving accounts of the applicants. Historic data indicate following distribution about the savings (in thousand dollars) of the applicant based on their classes.

$(x|0) \sim \text{Uniform}(35, 60)$, $(x|1) \sim \text{Uniform}(15, 45)$, $(x|2) \sim \text{Uniform}(0, 20)$

Furthermore, we know that 20% and 30% of the previous applicants are low-risk and medium-risk, respectively. If an applicant has 48,000 dollars in her savings account, predict the applicant's class (based on Bayes decision rules).

Hint: pdf for uniform distribution — $f(x) = \frac{1}{b-a}$ if $a < x < b$.

Answer:

$P(c_0|x) = 0.336$

$P(c_1|x) = 0.332$

$P(c_2|x) = 0.332$

$$P(c_0) = 0.2 \\ P(c_1) = 0.3 \\ P(c_2) = 0.5$$

$$P(C_0|x) = 0.2 \times \frac{1}{25} + 1 = \frac{1}{25} \\ P(C_1|x) = 0.3 \times 0 + 1 = 0 \\ P(C_2|x) = 0.5 \times 0 + 1 = 0 \\ P(X) = 0.2 \times \frac{1}{25} + 0 + 0 = \frac{1}{25}$$
$$P(c_0|x) = \frac{P(x|c_0)}{P(x|c_0) + P(x|c_1) + P(x|c_2)} = \frac{1}{3} \approx 0.333 \\ P(c_1|x) = \frac{1}{3} \approx 0.333 \\ P(c_2|x) = \frac{1}{3} \approx 0.333$$

3. (20 points) Let $f(x) = x^3 + 6x + 5$

3.1. Compute $f'(x)$, $f''(x)$, and determine whether f convex, concave, or neither.

Answer:

as we know:

- $f(x)$ is convex at x if $f''(x) > 0$
- $f(x)$ is concave at x if $f''(x) < 0$

$f'(x) = 3x^2 + 6$

$f''(x) > 0$, when $x > 0$; $f(x)$ is convex at x when $x > 0$

$f''(x) < 0$, when $x < 0$; $f(x)$ is concave at x when $x < 0$

3.2. Optimize the following function by finding the critical value(s) at which the function is optimized.

Hint: To find the critical values put $f'(x) = 0$

Answer:

critical values are at $f'(x) = 0$

$f'(x) = 3x^2 + 6 = 0$

$x^2 = -2$

$x = \pm i\sqrt{2}$

4. (20 points) Consider the following table as a training data for Naive Bayes classifier. Find the following probabilities. Apply Laplace correction if required.

A	B	C	LABEL
1	0	1	Yes
1	1	1	Yes
0	1	1	No
1	1	0	No
1	0	1	No

4.1 $P(\text{Label} = \text{Yes}|A = 1, B = 1, C = 0)$

Answer:

$P(\text{Yes}) = \frac{2}{6} = \frac{1}{3}$

$P(A=1) = \frac{3}{6} = \frac{1}{2}$

$P(B=1) = \frac{3}{6} = \frac{1}{2}$

$P(C=0) = \frac{4}{6} = \frac{2}{3}$

$P(\text{Yes}|A=1, B=1, C=0) = \frac{1}{3} \times \frac{1}{2} \times \frac{1}{2} \times \frac{2}{3} = \frac{1}{18}$

$P(\text{Yes}|A=1, B=1, C=0) = \frac{1}{18} \approx 5.56\%$

4.2 $P(\text{Label} = \text{No}|A = 1, B = 0, C = 0)$

Answer:

$P(\text{No}) = \frac{4}{6} = \frac{2}{3}$

$P(A=1) = \frac{3}{6} = \frac{1}{2}$

$P(B=0) = \frac{3}{6} = \frac{1}{2}$

$P(C=0) = \frac{4}{6} = \frac{2}{3}$

$P(\text{No}|A=1, B=0, C=0) = \frac{2}{3} \times \frac{1}{2} \times \frac{1}{2} \times \frac{2}{3} = \frac{2}{18} = \frac{1}{9}$

$P(\text{No}|A=1, B=0, C=0) = \frac{1}{9} \approx 11.11\%$

5. (20 points) Consider the pdf of $f(x)$ and calculate the following:

5. (20 points) Consider the pdf of $f(x)$ and calculate the following:

- find the value of c

- CDF of $f_X(x)$

- find $P(X \geq 1/2)$

$$f(x) = cx^3 \quad 0 \leq x \leq 1 \quad \text{(1)}$$

where c is a positive constant.

Required Formulas :

$$\frac{dx^n}{dx} = nx^{n-1}$$

$$\int x^n dx = \frac{x^{n+1}}{n+1}$$

5.1. find the value of c

Answer:

c>0 in order to maintain $f(x)>0$

let c be 1 as simply

5.2. CDF of $f(x)$

Answer:

Let c be 1, then CDF of $f(x)$ will be:

$F(x) = \int_0^x cx^3 dx = \frac{cx^4}{4}$

$F(x) = \frac{1}{4}x^4$

$F(1/2) = \frac{1}{4}(1/2)^4 = \frac{1}{64}$

$F(1/2) = 0.015625$

$F(1) = \frac{1}{4}(1)^4 = \frac{1}{4}$

$F(1) = 0.25$

$F(2) = \frac{1}{4}(2)^4 = \frac{16}{4} = 4$

$F(2) = 1$

$F(3) = \frac{1}{4}(3)^4 = \frac{81}{4} = 20.25$

$F(3) = 5.0625$

$F(4) = \frac{1}{4}(4)^4 = \frac{256}{4} = 64$

$F(4) = 16$

$F(5) = \frac{1}{4}(5)^4 = \frac{625}{4} = 156.25$

$F(5) = 39.0625$

$F(6) = \frac{1}{4}(6)^4 = \frac{1296}{4} = 324$

$F(6) = 81$

$F(7) = \frac{1}{4}(7)^4 = \frac{2401}{4} = 600.25$

$F(7) = 150.0625$

$F(8) = \frac{1}{4}(8)^4 = \frac{4096}{4} = 1024$

$F(8) = 256$

$F(9) = \frac{1}{4}(9)^4 = \frac{6561}{4} = 1640.25$

$F(9) = 410.0625$

$F(10) = \frac{1}{4}(10)^4 = \frac{10000}{4} = 2560$

$F(10) = 640$

$F(11) = \frac{1}{4}(11)^4 = \frac{14641}{4} = 3660.25$

$F(11) = 915.0625$

$F(12) = \frac{1}{4}(12)^4 = \frac{20736}{4} = 5184$

$F(12) = 1296$

$F(13) = \frac{1}{4}(13)^4 = \frac{28561}{4} = 7140.25$

$F(13) = 1785.0625$

$F(14) = \frac{1}{4}(14)^4 = \frac{36904}{4} = 9226$

$F(14) = 2306.5$

$F(15) = \frac{1}{4}(15)^4 = \frac{50625}{4} = 12656.25$

$F(15) = 3164.0625$

$F(16) = \frac{1}{4}(16)^4 = \frac{65536}{4} = 16384$

$F(16) = 4096$

$F(17) = \frac{1}{4}(17)^4 = \frac{83521}{4} = 20880.25$

$F(17) = 5220.0625$

$F(18) = \frac{1}{4}(18)^4 = \frac{104976}{4} = 26244$