[C] [R] [L]

بسم الله الرحمن الرحیم

دانشکده مهندسی برق و کامپیوتر

یادگیری ماشین - تمرین دوم

سید مهدی رضوی

استاد : آقای دکتر توسلی پور - آقای دکتر ابوالقاسمی

سید مهدی رضوی

اسفند ماه ۱۴۰۲

فهرست مطالب

فهرست تصاویر

۱ تمرین اول

واریانس تک تک این n متغیر تصادفی برابر با مجموع واریانس آن‌ها خواهد بود.
می‌دانیم واریانس متغیر تصادفی برابر با امیدریاضی متغیر منهای میانگین آن متغیر خواهد بود.

$$\sigma_n^2(x) = \sum_{i=1}^{n} \varepsilon \left[\left(\frac{1}{nV_n} \phi \left(\frac{x - x_i}{h_n} - \frac{1}{n} \overline{P_n}(x) \right) \right)^2 \right]$$
$$\sigma_n^2(x) = n\varepsilon \left[\tfrac{1}{n^2 V_n^2} \phi^2 \left(\tfrac{x - x_i}{h_n} \right) \right] - \tfrac{1}{n} \bar{P}_n^2(x)$$
$$\sigma_n^2(x) = \frac{1}{nV_n} \int \frac{1}{V_n} \phi^2 \left(\frac{x - V}{h_n} \right) P(V) dV - \frac{1}{n} \bar{P}_n^2(x)$$

برای بدست آوردن کوچکترین کران بالا واریانس متغیر ایکس ، می‌توانیم عبارت $-\frac{1}{n}\bar{P}_n^2(x)$ در نظرنگیریم ، چون که واریانس متغیر ایکس همواره از عبارت سمت چپ کوچکتر خواهدبود.
همانطور که روابط زیر را می‌دانیم :

$$P_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{V_n} \phi \left(\frac{x - x_i}{h_n} \right)$$

$$\sigma_n^2(x) <= \frac{\sup(\phi) \overline{P_n}(x)}{nV_n}$$

۲ تمرین دوم

Given the expression $p(x) = \frac{1}{a}\left(1 - e^{-\frac{x}{h_n}}\right)$ for $0 < x < a$ and $p(x) = \frac{1}{a}\left(e^{\frac{a}{h_n}} - 1\right)e^{-\frac{x}{h_n}}$ for $x \geq a$, we aim to calculate the mean estimate of the Parzen window function using this probability distribution.

The mean estimate of a function $g(x)$ with respect to a probability distribution $p(x)$ over the interval $[0, a]$ is given by:

$$\text{Mean}(g) = \int_0^a g(x) \cdot p(x)\, dx$$

Given the probability distribution $p(x)$ as specified, and the Parzen window function $\varphi(x) = e^{-x}$ if $x > 0$ else 0, we can calculate the mean estimate as follows:

$$\text{Mean}(\varphi) = \int_0^a \varphi(x) \cdot p(x)\, dx = \int_0^a e^{-x} \cdot \frac{1}{a}\left(1 - e^{-\frac{x}{h_n}}\right) dx$$

Now, we need to split the integral into two parts: one for $0 < x < a$ and another for $x \geq a$.

For $0 < x < a$:

$$\begin{aligned}
&\int_0^a e^{-x} \cdot \frac{1}{a}\left(1 - e^{-\frac{x}{h_n}}\right) dx \\
&= \frac{1}{a}\int_0^a e^{-x} - e^{-x-\frac{x}{h_n}}\, dx \\
&= \frac{1}{a}\left[-e^{-x} + h_n e^{-x-\frac{x}{h_n}}\right]\Bigg|_0^a \\
&= \frac{1}{a}\left[-e^{-a} + h_n e^{-a-\frac{a}{h_n}} + e^0 - h_n e^0\right]
\end{aligned}$$

:$x \geq a$ For

$$\begin{aligned}
&\int_a^\infty e^{-x} \cdot \frac{1}{a}\left(e^{\frac{a}{h_n}} - 1\right)e^{-\frac{x}{h_n}}\, dx \\
&= \frac{1}{a}\left(e^{\frac{a}{h_n}} - 1\right)\int_a^\infty e^{-\frac{2x}{h_n}+\frac{a}{h_n}}\, dx \\
&= \frac{1}{a}\left(e^{\frac{a}{h_n}} - 1\right)\left[\frac{h_n}{2h_n}e^{\frac{a}{h_n}-\frac{2a}{h_n}}\right] \\
&= \frac{1}{a}\left(e^{\frac{a}{h_n}} - 1\right)\cdot\frac{1}{2}e^{-\frac{a}{h_n}}
\end{aligned}$$

Therefore, the mean estimate of the Parzen window function with respect to the given probability distribution is the sum of these two parts:

$$\begin{aligned}
\text{Mean}(\varphi) = &\frac{1}{a}\left[-e^{-a} + h_n e^{-a-\frac{a}{h_n}} + e^0 - h_n e^0\right] \\
&+ \frac{1}{a}\left(e^{\frac{a}{h_n}} - 1\right)\cdot\frac{1}{2}e^{-\frac{a}{h_n}}
\end{aligned}$$

This expression represents the mean estimate of the Parzen window function based on the specified probability distribution. Further simplification can be done if necessary.

۳ تمرین سوم

we regression، L۲ in function objective (RLS) Squares Least Regularized the to solution the find To function: following the minimize to need

$$\text{minimize:} Y - X\beta^2 + \lambda\beta^2$$

β – variables، predictor of matrix design the is X – values، target observed of vector the is Y – where: parameter. regularization the is λ – estimated، be to coefficients of vector the is
function objective The .$R(\beta)$ as $\lambda\beta^2$ term second the and $L(\beta)$ as $Y - X\beta^2$ term first the denote Let's as: rewritten be can

$$\text{minimize:} L(\beta) + R(\beta)$$

$$\nabla(L(\beta) + R(\beta)) = 0$$

get: we operation، gradient the Expanding

$$\nabla L(\beta) + \nabla R(\beta) = 0$$

us: gives β to respect with $L(\beta)$ of derivative the Taking

$$\nabla L(\beta) = -2X^T(Y - X\beta)$$

us: gives β to respect with $R(\beta)$ of derivative the Taking

$$\nabla R(\beta) = 2\lambda\beta$$

as: equation the rewrite can we Now،

$$-2X^T(Y - X\beta) + 2\lambda\beta = 0$$

get: we terms، the Rearranging

$$X^T X\beta - X^T Y + \lambda\beta = 0$$

$$(X^T X + \lambda I)\beta = X^T Y$$

$$\beta = (X^T X + \lambda I)^{-1} X^T Y$$

۴ تمرین چهارم

۵ تمرین پنجم

To determine the Bayesian decision rule and the error of classification based on the given class-conditional probability distributions, we can use the Bayes' theorem and the minimum error rate criterion.

distribution probability class–conditional the with w_2 Class –

.۰ else $\frac{1}{3} < x < 1$ if $P(x|w_2) = \frac{3}{2}$

The Bayesian decision rule states that we should classify an observation x to the class that maximizes the posterior probability, which can be calculated using Bayes' theorem:

$$P(w_i|x) = \frac{P(x|w_i) \cdot P(w_i)}{P(x)}$$

class– the is $P(x|w_i)$ – .x observation given w_i class of probability posterior the is $P(w_i|x)$ – where: is $P(x)$ – .w_i class of probability prior the is $P(w_i)$ – .w_i class of distribution probability conditional .$P(x) = \sum_i P(x|w_i) \cdot P(w_i)$ as calculated be can which ،x observation of probability total the

The error of classification can be calculated as the probability of misclassification, which can be expressed as:

$$P(\text{error}) = \int_{-\infty}^{\infty} \min_i [P(w_i|x)] \cdot P(x)\, dx$$

In this specific case with the given class-conditional probability distributions and assuming equal prior probabilities (i.e., $P(w_1) = P(w_2) = \frac{1}{2}$), we can apply the Bayesian decision rule to determine the optimal decision boundary and calculate the error of classification based on the provided information.

class. each for set training the in neighbor nearest the find ،x point test a Given .۱

neighbor. nearest its of class the to x point test the Assign .۲

labels. class correct their and points test possible all considering by probability error the Calculate .۳

follows: as calculated be can classifier nearest–neighbor the of probability error The
Let:

- $w1$ class of probability prior the be $P(w1)$
- $w2$ class of probability prior the be $P(w2)$
- $w1$ class given x observing of function density probability conditional the be $P(x|w1)$
- $w2$ class given x observing of function density probability conditional the be $P(x|w2)$
- $x2$ and $x1$ points two between metric distance the be $d(x1, x2)$

by: given is classifier nearest–neighbor the using classification binary for probability error The

$$P(error) = P(w1) \times P(error|w1) + P(w2) \times P(error|w2)$$

Where:

- $P(error|w1) = \int_x (P(w2|x) \times P(x|w1))dx$
- $P(error|w2) = \int_x (P(w1|x) \times P(x|w2))dx$