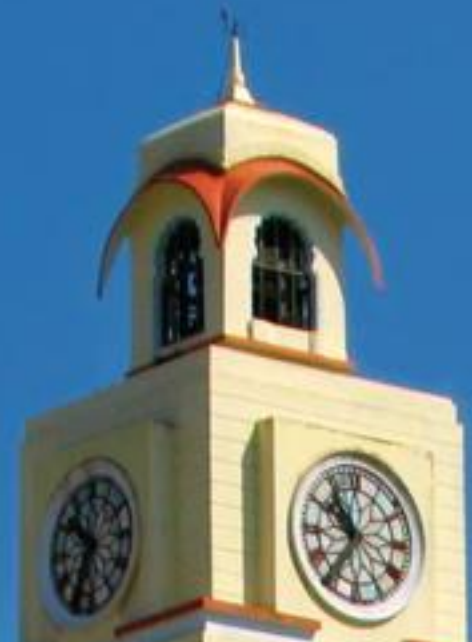


Machine Learning

CS F464

BITS Pilani
Dubai Campus

Dr. Pranav M. Pawar

Contents

- Gaussian for classification

Gaussian for Classification

- Used for classifying continuous data.
- Consider a random variable x whose probability density function (pdf), given class *C* is a Gaussian with mean μ_c and variance σ_c^2. Using Bayes' theorem,

$$
\begin{aligned}
P(C \mid x) &\propto p(x \mid C)\, P(C) \\
&\propto N(x; \mu_c, \sigma_c^2)\, P(C) \\
&\propto \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(\frac{-(x-\mu_c)^2}{2\sigma_c^2} \right) P(C)
\end{aligned}
$$

- Where, $p(x/C)$ is the likelihood of class *C* given observation *x*.
- $\mu_c = \frac{\sum_{i=1}^{n} x_i}{n}$, $\sigma_c^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_c)^2$

Log Likelihood and Log Probabilities (1)

- Use log to deal with Gaussian.
- Log likelihood of the Gaussian probability distribution is, $LL(x|C) = LL(x|\mu_c, \sigma_c^2)$

$$
\begin{aligned}
LL(x|\mu_c, \sigma_c^2) &= \ln p(x|\mu_c, \sigma_c^2) \\
&= \ln\left[\frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(\frac{-(x-\mu_c)^2}{2\sigma_c^2}\right)\right] \\
&= -\ln\left(\sqrt{2\pi\sigma_c^2}\right) - \frac{(x-\mu_c)^2}{2\sigma_c^2} \\
&= \frac{1}{2}\left(-\ln(2\pi) - \ln\sigma_c^2 - \frac{(x-\mu_c)^2}{\sigma_c^2}\right)
\end{aligned}
$$

Log Likelihood and Log Probabilities (2)

- Using Bayes theorem Log posterior probabilities are,

$$LP(C|x) = LL(x|C) + LP(C) + \text{const.}$$
$$= \frac{1}{2}\left(-\ln(2\pi) - \ln\sigma_c^2 - \frac{(x-\mu_c)^2}{\sigma_c^2}\right) + \ln P(C) + \text{const.}$$

- where *LP(C)* is the log prior probability of class *C*, and '*const.*' is the log of the constant of proportionality, *p(x).*

Log probability ratio

- If $c1$ and $c2$ are modelled by Gaussians with means μ_a and μ_b, and variances σ_a^2 and σ_b^2 , then we can write the log odds (ratio of posterior probabilities) as follows,

$$\ln \frac{P(c_1 | x)}{P(c_2 | x)} = \ln P(c_1 | x) - P(c_2 | x)$$

$$\begin{aligned}\ln \frac{P(c_1 | x)}{P(c_2 | x)} &= \frac{1}{2}\left(-\ln(2\pi) - \ln\sigma_a^2 - \frac{(x-\mu_a)^2}{\sigma_a^2}\right) \\ &\quad - \frac{1}{2}\left(-\ln(2\pi) - \ln\sigma_b^2 - \frac{(x-\mu_b)^2}{\sigma_b^2}\right) + (\ln P(c_1) - \ln P(c_2)) \\ &= -\frac{1}{2}\left(\frac{(x-\mu_a)^2}{\sigma_a^2} - \frac{(x-\mu_b)^2}{\sigma_b^2} + \ln\sigma_a^2 - \ln\sigma_b^2\right) + \ln P(c_1) - \ln P(c_2) \,. \qquad (4\end{aligned}$$

Example (1)

- Consider a following dataset,

Class S	10	8	10	10	11	11
Class T	12	9	15	10	13	13

  - The following unlabelled data points are available: $x1 = 10$, $x2 = 11$, $x3 = 6$
  - To which class each data points will be assigned if the two classes have equal prior probabilities.

- Step 1: Find the mean and variance

$$\hat{\mu}(S) = 10 \qquad \hat{\sigma}^2(S) = 1$$

$$\hat{\mu}(T) = 12 \qquad \hat{\sigma}^2(T) = 4$$

Example (2)

Step 2: Calculate log posterior probability ratio.

$$\ln \frac{P(S \mid X=x)}{P(T \mid X=x)} = -\frac{1}{2}\left(\frac{(x-\mu_S)^2}{\sigma_S^2} - \frac{(x-\mu_T)^2}{\sigma_T^2} + \ln \sigma_S^2 - \ln \sigma_T^2\right) + \ln P(S) - \ln P(T)$$

$$= -\frac{1}{2}\left(\frac{(x-\mu_S)^2}{\sigma_S^2} - \frac{(x-\mu_T)^2}{\sigma_T^2} + \ln \sigma_S^2 - \ln \sigma_T^2\right)$$

$$= -\frac{1}{2}\left((x-10)^2 - \frac{(x-12)^2}{4} - \ln 4\right)$$

If the log ratio is less than 0, then assign to class *T*, otherwise assign to class *S* .

Example (3)

Step 3: Consider each value of x and calculate log posterior probability ratio.

$x_1 = 10$:

$$\ln \frac{P(S \mid X = x_1)}{P(T \mid X = x_1)} = -\frac{1}{2}\left((x_1 - 10)^2 - \frac{(x_1 - 12)^2}{4} - \ln 4\right)$$
$$= -\frac{1}{2}(0 - 1 - \ln 4)$$
$$= 1.19$$

Class S

$x_2 = 11$:

$$\ln \frac{P(S \mid X = x_2)}{P(T \mid X = x_2)} = -\frac{1}{2}\left((x_2 - 10)^2 - \frac{(x_2 - 12)^2}{4} - \ln 4\right)$$
$$= -\frac{1}{2}(1 - 0.25 - \ln 4)$$
$$= 0.32$$

Class S

Example (4)

$x_3 = 6$:

$$\ln \frac{P(S \mid X=x_3)}{P(T \mid X=x_3)} = -\frac{1}{2}\left((x_3-10)^2 - \frac{(x_3-12)^2}{4} - \ln 4\right)$$
$$= -\frac{1}{2}(16 - 9 - \ln 4)$$
$$= -2.81$$

Class T

- Practice Problem:

Consider different probabilities, P (S) = 0.3 and P(T) = 0.7. Including this prior information, to which class should each of the above test data points {*x*1, *x*2, *x*3} be assigned?

Hint: The log posterior probability ratios is,

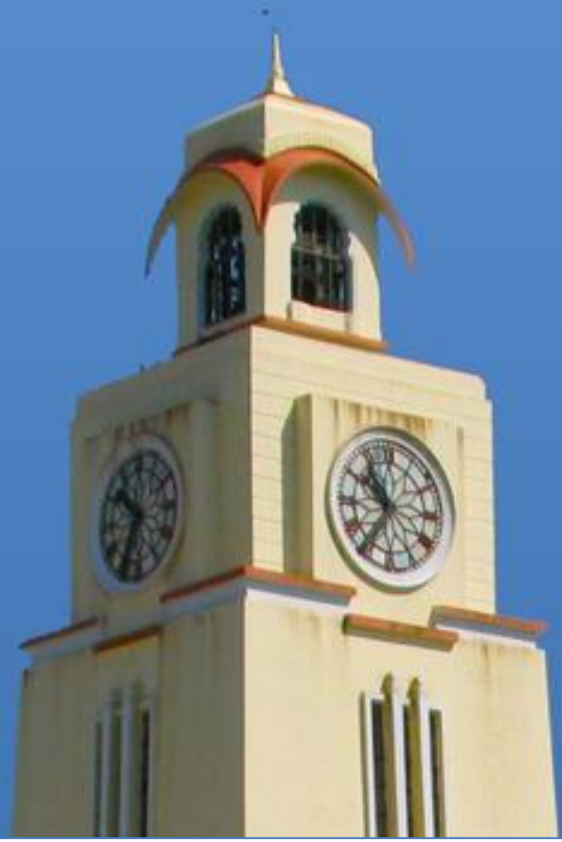
$$\ln \frac{P(S \mid X=x)}{P(T \mid X=x)} = -\frac{1}{2}\left(\frac{(x-\mu_S)^2}{\sigma_S^2} - \frac{(x-\mu_T)^2}{\sigma_T^2} + \ln \sigma_S^2 - \ln \sigma_T^2\right) + \ln P(S) - \ln P(T)$$
$$= -\frac{1}{2}\left((x-10)^2 - \frac{(x-12)^2}{4} - \ln 4\right) + \ln P(S) - \ln P(T)$$
$$= -\frac{1}{2}\left((x-10)^2 - \frac{(x-12)^2}{4} - \ln 4\right) + \ln(3/7)$$

Sources

- http://people.cs.ksu.edu/~hankley/d764/Slides07/Chou_Reasoning.ppt
- http://www.inf.ed.ac.uk/teaching/courses/inf2b/learnnotes/
- Chapter 6, Tom M. Mitchell, Machine Learning, The McGraw-Hill Companies, 1st edition 2013.
- https://www.udemy.com/course/python-for-data-science-and-machine-learning-bootcamp

innovate achieve lead

BITS Pilani
Dubai Campus



Thank You!