Homework 2: Classification

TDA231 - Algorithms for Machine Learning & Inference

Theodor Åstrand, theast@student.chalmers.se, 931109-9114
Linnéa Otterlind, linott@student.chalmers.se, 921126-3620

Chalmers University of Technology

February, 2016

Problem 1.1

(a)

"not rich", "married" and "healthy" translates to $(0,1,1)$. From naive Bayes we get:

$P(\mathbf{x}_{new}|c_{new}=k,\mathbf{X},\mathbf{c}) = \prod_{d=1}^{D} P(x_d^{new}|c_{new}=k,\mathbf{X},\mathbf{c})$

Which for our problem translates to:

$P((0,1,1)|c_{new}=1,\mathbf{X},\mathbf{c}) = \prod_{d=1}^{3} P(x_d^{new}|c_{new}=1,\mathbf{X},\mathbf{c}) =$

$= P(x_1^{new}=0|c_{new}=1,\mathbf{X},\mathbf{c}) \cdot P(x_2^{new}=1|c_{new}=1,\mathbf{X},\mathbf{c}) \cdot P(x_3^{new}=1|c_{new}=1,\mathbf{X},\mathbf{c}) =$

$= \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{3}{4} = \frac{3}{32}$

$P((0,1,1)|c_{new}=0,\mathbf{X},\mathbf{c}) = \prod_{d=1}^{3} P(x_d^{new}|c_{new}=0,\mathbf{X},\mathbf{c}) =$

$= P(x_1^{new}=0|c_{new}=0,\mathbf{X},\mathbf{c}) \cdot P(x_2^{new}=1|c_{new}=0,\mathbf{X},\mathbf{c}) \cdot P(x_3^{new}=1|c_{new}=0,\mathbf{X},\mathbf{c}) =$

$= \frac{3}{4} \cdot \frac{1}{4} \cdot \frac{1}{4} = \frac{3}{64}$

$P(c_{new}=1) = P(c_{new}=0) = \frac{1}{2}$

$P(c_{new}=1|(0,1,1),\mathbf{X},\mathbf{c}) = \frac{P((0,1,1)|c_{new}=1,\mathbf{X},\mathbf{c}) \cdot P(c_{new}=1)}{\sum_{i=0}^{1} P((0,1,1)|c_{new}=i,\mathbf{X},\mathbf{c}) \cdot P(c_{new}=i)} = \frac{\frac{3}{32} \cdot \frac{1}{2}}{\frac{3}{32} \cdot \frac{1}{2} + \frac{3}{64} \cdot \frac{1}{2}} = \frac{2}{3}$

(b)

$P((0,1)|c_{new}=1,\mathbf{X},\mathbf{c}) = \prod_{d=1}^{2} P(x_d^{new}|c_{new}=1,\mathbf{X},\mathbf{c}) =$

$= P(x_1^{new}=0|c_{new}=1,\mathbf{X},\mathbf{c}) \cdot P(x_2^{new}=1|c_{new}=1,\mathbf{X},\mathbf{c}) = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$

$P((0,1)|c_{new}=0,\mathbf{X},\mathbf{c}) = \prod_{d=1}^{2} P(x_d^{new}|c_{new}=0,\mathbf{X},\mathbf{c}) =$

$= P(x_1^{new}=0|c_{new}=0,\mathbf{X},\mathbf{c}) \cdot P(x_2^{new}=1|c_{new}=0,\mathbf{X},\mathbf{c}) = \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}$

$P(c_{new}=1|(0,1),\mathbf{X},\mathbf{c}) = \frac{P((0,1)|c_{new}=1,\mathbf{X},\mathbf{c}) \cdot P(c_{new}=1)}{\sum_{i=0}^{1} P((0,1)|c_{new}=i,\mathbf{X},\mathbf{c}) \cdot P(c_{new}=i)} =$

$= \frac{\frac{1}{8} \cdot \frac{1}{2}}{\frac{1}{8} \cdot \frac{1}{2} + \frac{3}{16} \cdot \frac{1}{2}} = \frac{2}{5} = 0.4$

Problem 1.2

There will be difficulties training Naive Bayes using this approach because x_1, x_2 and x_3 are not independent, exactly one of them should be one at any given time. Naive Bayes requires that the dimensions of $\mathbf{x}$ are independent within each class. We could extend Naive Bayes to only use one variable that could vary between three given values, each value representing each of the three first dimensions of the original $\mathbf{x}$.

$$\mathrm{x}_1 = \begin{cases} 0 \text{ if customer is younger than 20.} \\ 1 \text{ if customer is between 20 and 30 in age.} \\ 2 \text{ if customer is older than 30.} \end{cases}$$

$x_2 = 1$ if customer walks to work and 0 otherwise.

Problem 2.1

(a)

$P(y_{new} = 1) = P(y_{new} = -1) = \frac{1}{2}$

$P1 = P(y_{new} = 1|\mathbf{x}_{new}, \mathbf{X}, \mathbf{y}) = \frac{P(x_{new}|y_{new}=1,\mathbf{X},\mathbf{y})\cdot\frac{1}{2}}{P(x_{new}|y_{new}=1,\mathbf{X},\mathbf{y})\cdot\frac{1}{2}+P(x_{new}|y_{new}=-1,\mathbf{X},\mathbf{y})\cdot\frac{1}{2}} =$

$= \frac{P(\mathbf{x}_{new}|\hat{\mu}_1,\hat{\sigma}_1^2)}{P(\mathbf{x}_{new}|\hat{\mu}_1,\hat{\sigma}_1^2)+P(\mathbf{x}_{new}|\hat{\mu}_2,\hat{\sigma}_2^2)}$

$P2 = P(y_{new} = -1|\mathbf{x}_{new}, \mathbf{X}, \mathbf{y}) = \frac{P(x_{new}|y_{new}=-1,\mathbf{X},\mathbf{y})\cdot\frac{1}{2}}{P(x_{new}|y_{new}=1,\mathbf{X},\mathbf{y})\cdot\frac{1}{2}+P(x_{new}|y_{new}=-1,\mathbf{X},\mathbf{y})\cdot\frac{1}{2}} =$

$= \frac{P(\mathbf{x}_{new}|\hat{\mu}_2,\hat{\sigma}_2^2)}{P(\mathbf{x}_{new}|\hat{\mu}_1,\hat{\sigma}_1^2)+P(\mathbf{x}_{new}|\hat{\mu}_2,\hat{\sigma}_2^2)}$

$P(\mathbf{x}_{new}|\hat{\mu}_1, \hat{\sigma}_1^2) = \frac{1}{(2\pi)^{3/2}\hat{\sigma}_1^3}\exp\big(-\frac{1}{2\hat{\sigma}_1^2}(\mathbf{x}-\hat{\mu}_1)^T(\mathbf{x}-\hat{\mu}_1)\big)$

$P(\mathbf{x}_{new}|\hat{\mu}_2, \hat{\sigma}_2^2) = \frac{1}{(2\pi)^{3/2}\hat{\sigma}_2^3}\exp\big(-\frac{1}{2\hat{\sigma}_2^2}(\mathbf{x}-\hat{\mu}_2)^T(\mathbf{x}-\hat{\mu}_2)\big)$

(d)

Test	sph_bayes	new_classifier
1	0.0200	0.0100
2	0.0175	0.0175
3	0.0100	0.0025
4	0.0100	0.0100
5	0.0100	0.0025
Mean	0.0135	0.0085

Problem 2.2

(c)

We can clearly see that using the feature function makes classification worse. We also noticed that the feature function took a really long time to evaluate.

Test	new_classifier	new_classifier (feature function)
1	0.0818	0.2523
2	0.0795	0.2568
3	0.0727	0.2205
4	0.0841	0.2341
5	0.0818	0.2136
Mean	0.0800	0.2355