Machine Learning, 2021 Fall
Assignment 2

Notice

Due 23:59 (CST), Nov 9, 2021

Plagiarizer will get 0 points.

LaTeXis highly recommended. Otherwise you should write as legibly as possible.

1 Naïve Bayes

Given the training data in the table below, predict the class of the following new example using Naïve Bayes classification, the feature is [rainy,mudium,true,low].

RID	outlook	temperature	windy	humidity	play
1	rainy	high	false	low	no
2	rainy	high	false	high	no
3	overcast	high	false	low	yes
4	sunny	medium	false	low	yes
5	sunny	low	true	low	yes
6	sunny	low	true	high	no
7	overcast	low	true	high	yes
8	rainy	medium	false	low	no
9	rainy	low	true	low	yes
10	sunny	medium	true	low	yes
11	rainy	medium	true	high	yes
12	overcast	medium	false	high	yes
13	overcast	high	ture	low	yes
14	sunny	medium	false	low	no

E = outlook: rainy, temperature:medium, windy:true, humidity: low

E_1 is outlook:rainy, E_2 is temperature:medium, E_3 is windy:true, E_4 is humidity: low.

We need to compute P(yes | E) and P(no | E) and compare them.

$P(\text{ yes } \mid E)=\frac{P(E_1 \mid \text{ yes }) P(E_2 \mid \text{ yes }) P(E_3 \mid \text{ yes }) P(E_4 \mid \text{ yes }) P(\text{ yes })}{P(E)}$

$P(yes \mid E)=\frac{0.2220*0.4440*0.6670*0.6680*0.443}{P(E)}=\frac{0.028}{P(E)}$

$P(no \mid E)=\frac{0.60*0.40*0.20*0.60*0.357}{P(E)}=\frac{0.010}{P(E)}$

Hence, the Naïve Bayes classifier predicts yes for the new example.

Suppose that Y follows a binomial distribution with parameters n and $p = \theta$, so that the p.m.f. of Y given θ is:

$$g(y|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}$$

for $y = 0, 1, \ldots, n$. Suppose that the prior p.d.f. of the parameter θ is the beta p.d.f., that is:

$$h(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

for $0 < \theta < 1$. Find the posterior p.d.f of θ, given that $Y = y$. That is, find $k(\theta \mid y)$.

hint: $\int_0^1 \left(\frac{\Gamma(y+\alpha+n-y+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\right) \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} d\theta = 1$

First, we find the joint p.d.f. of the statistic Y and the parameter θ by multiplying the prior p.d.f. $h(\theta)$ and the conditional p.m.f. of Y given θ. That is:

$$k(y,\theta) = g(y \mid \theta)h(\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

over the support $y = 0, 1, 2, \ldots, n$ and $0 < \theta < 1$. Simplifying by collecting like terms, we get that the joint p.d.f. of the statistic Y and the parameter θ is:

$$k(y,\theta) = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

over the support $y = 0, 1, 2, \ldots, n$ and $0 < \theta < 1$. Then, we find the marginal p.d.f. of Y by integrating $k(y, \theta)$ over the parameter space of θ :

$$k_1(y) = \int_0^1 k(y,\theta) d\theta = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} d\theta$$

Now, if we multiply the integrand by 1 in a special way:

$$k_1(y) = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(y+\alpha+n-y+\beta)}\right) \int_0^1 \left(\frac{\Gamma(y+\alpha+n-y+\beta)}{\Gamma(y+\alpha)\Gamma(n-y+\beta)}\right) \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1} d\theta$$

we see that we get a beta p.d.f. with parameters $y + \alpha$ and $n - y + \beta$ that therefore, by the definition of a valid p.d.f., must integrate to 1. Simplifying we therefore get that the marginal p.d.f. of Y is:

$$k_1(y) = \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(y+\alpha+n-y+\beta)}\right)$$

on the support $y = 0, 1, 2, \ldots, n$. Then, the posterior p.d.f. of θ, given that $Y = y$ is:

$$k(\theta|y) = \frac{k(y,\theta)}{k_1(y)} = \frac{\binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}{}^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}}{\binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \left(\frac{\Gamma(y+\alpha)\Gamma(n-y+\beta)}{\Gamma(y+\alpha+n-y+\beta)}\right)}$$

Because some things cancel out, we see that the posterior p.d.f. of θ, given that $Y = y$ is:

$$k(\theta|y) = \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+y)\Gamma(n+\beta-y)} \theta^{y+\alpha-1}(1-\theta)^{n-y+\beta-1}$$

for $0 < \theta < 1$, which you might recognize as a beta p.d.f. with parameters $y + \alpha$ and $n - y + \beta$.