531 HW 3

Peter Buonaiuto

October 2023

1 Bayes

Question Classify the point (Age=23, Car=truck) via the full and naive Bayes approach, assuming that the domain of Car is given as sports, vintage, suv, truck

x_i	Age	Car	Class
x_1	25	sports	L
x_2	20	vintage	H
x_3	25	sports	L
x_4	45	suv	H
x_5	20	sports	H
x_6	25	suv	H

Figure 1: Training Data

1.a Naive Bayes

To compute by naive bayes, I will assume all attributes to be independent and thus compute the variances, yet assume the covariances to be 0 in the covariance matrix. Additionally, I will apply smoothing by adding the one count of each data type to avoid 0 values, while maintaining the relational conditions between all types in effort to prevent bias towards the category of 0 training presence. Finally, due to the assumption of indepence, I can compute my posterior probabilities by the product of the probability for each independent attribute with the prior probability, and then normalize.

Compute Priors

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

$$P(H) = \frac{4}{6} = \frac{2}{3} \tag{2}$$

Compute Age Likelihood for class L asuming age to be categorical from 20 to 45 (26 values)
Since age 23 does not appear in the dataset, I use smoothing to introduce a small probability for the point. I do so by adding 1 to the numerator to simulate the value being in the dataset, and then increase the denominator by the range of the feature, 23, to offset the influence of the pseudo point.
$P(age = 23|L) = \frac{0+1}{2+26} = \frac{1}{28}$

Likelihood of age for class H In a similar fashion, I will find the likelihood of P(age = 23 | H)

$P(age = 23|H) = \frac{0+1}{4+26} = \frac{1}{30}$

Computing Likelihood of Truck using laplace smoothing
P(type = truck | Class L) = $\frac{0+1}{2+4} = \frac{1}{6}$
P(type = truck | Class H) = $\frac{0+1}{4+4} = \frac{1}{8}$

Computing posterior probabilities by multiplying prior by the calculated value above, and then normalizing
P(X | L) = $P(L) * P(truck|L) * P(23|L) = \frac{1}{3} * \frac{1}{6} * \frac{1}{28} = 0.0020$
P(X | H) = $P(H) * P(truck|H) * P(23|H) = \frac{2}{3} * \frac{1}{8} * \frac{1}{30} = 0.0028$

Normalization We now normalize the posterior probabilities so that they sum to one. We do this by representing each as a part with respect to all classes.
P(X | L)' = $\frac{P(X|L)}{P(X|L)+P(X|H)} = \frac{0.0020}{0.0048} = 0.417$
P(X | H)' = $\frac{P(X|H)}{P(X|L)+P(X|H)} = \frac{0.0028}{0.0048} = 0.583$

Classification We therefore classify the new data point (Age = 23, Type = Truck) as class H since class H maximizes P(X $|c_i$)

1.b Full Bayes

We now release the assumption of independence, so can not find the independent probabilities of the features and then multiple. We must find the join probability. However, since the datapoint does is not yet classified, we must use smoothing to introduce a small probability to such a point to avoid 0s which would cause our normalization to be undefined, preventing a classification.

In the case of full bayes, we smooth by increasing the denominator by the product of the ranges of the features, as opposed to in neive where we only only consider one feature at a time. We now consider both features for the reason of finding joint probabilities, leading to the reason for the product in the denominator.

Compute Priors

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

$$P(H) = \frac{4}{6} = \frac{2}{3} \tag{4}$$

Likelihood of the stated point given L We compute the joint probabilities using the above definition and reason for smoothing.
$P(age = 23, type = truck|L) = \frac{0+1}{2+(26*4)} = \frac{1}{106}$

Likelihood of the stated point given H Similiarly for H, but the denominator includes the product plus the number of items in class H
$P(age = 23, type = truck|H) = \frac{0+1}{4+(26*4)} = \frac{1}{108}$

Computing posterior probabilities by multiplying prior by the calculated value above, and then normalizing
P(X | L) = $P(L) * P(truck, 23|L) = \frac{1}{3} * \frac{1}{106} = 0.0031$
P(X | H) = $P(H) * P(truck, 23|H) = \frac{2}{3} * \frac{1}{108} = 0.0062$

Normalization We now normalize the posterior probabilities so that they sum to one. We do this by representing each as a part with respect to all classes, by dividing the value by the sum of all to represent it as a part of the whole.
P(X | L)' = $\frac{P(X|L)}{P(X|L)+P(X|H)} = \frac{0.0031}{0.0093} = 0.333$
P(X | H)' = $\frac{P(X|H)}{P(X|L)+P(X|H)} = \frac{0.0062}{0.0093} = 0.667$

Classification We therefore classify the new data point (Age = 23, Type = Truck) as class H since class H maximizes P(X $|c_i$)

2 Logistic Regression

2.d Cross Entropy

Observations In general, the cross entropy loss seems to decrease as the epochs rise. There is some fluctuation, but almost always, the final epoch's cross entropy loss is below that of the first epoch.

With respect to the average cross entropies, that of the test data is lower than that of the training data. This suggests that the training data formed a good model for w since we experience lower error with the testing data.