Name: Shreeyash S. Dongarkar

PRN: 22510025

Machine Learning Lab Assignment 4

Study of Multivariate Classifiers

We can calculate posterior probability in case of a binary classifier to make classifications. It is probability of a class given observed features and is denoted as

Using Bayes’ Theorem, it is computed as:

This tells us how likely is given the observed data.

Since the denominator is same for both the classes, classification depends on comparing:

Case 1: When features are not correlated

We know that,

As we are ignoring the denominator, we can write by **chain rule of probability**

Here are conditionally independent hence

Since multiplying with many probabilities can lead to very small values let’s take *log* on both sides

As is constant for all classes, classification rule simplifies to:

Now we have two not corelated input features:

= Height (in cm) and = Haemoglobin Level in (g/dL)

And we have two classes male and female

Here we assume that each feature follows Gaussian Distribution within each class.

Hence the decision rule becomes:

where,

Is the probability density function which tells the likelihood of observing given feature under a Gaussian Distribution.

Hence the final decision rule becomes:

Observations:

* Higher prior probability favours class
* Closer values to the mean increase the likelihood
* Smaller variance gives the feature more weight in classification

Case 2: When features are correlated

We know,

Taking log and ignoring the denominator

For multivariate gaussian distribution

where,

* = Feature Vector (Vector containing all n features)
* = Mean Vector (Mean vector representing average of each feature for that class)
* = Covariance Matrix
* represents the Mahalanobis Distance which measures how far a point is from the class mean , accounting for correlation

Ignoring constant terms

Final decision rule becomes,

Now we have two not corelated input features:

= Height (in cm) and = Haemoglobin Level in (g/dL)

And we have two classes male and female

Plugging these values we get,

Observations:

* When features are correlated appears, adjusting for correlation which means both features together influence classification.
* When features are not correlated and each feature contributes independently.
* High covariance implies diagonal decision boundary as a change in height also strongly implies a change in weight
* And low covariance the boundary is more aligned with feature axes.

p-Values can also help us in determining whether a feature is statistically significant for classification. A low p-value suggests that feature is useful. Hence for handling correlated features we can check correlation between the features and perform a statistical test to find p-value of each feature and remove the features with high p-values.

Also, here as height and weight are correlated, we can derive a feature from them rather than using both independently as Body Mass Index (BMI)

* This removes the dependency between the both and thus reducing dimensionality
* But this might not be useful in each case as transforming may not be useful and may not improve accuracy.

How far can accuracy be improved by adding features?

We can start with a base model consisting of two features and add one feature at a time and measure accuracy. We should stop when improvement in accuracy is negligible. We can divide the dataset into multiple subsets to evaluate how well the model generalizes to unseen data. This helps ensure that the model performs consistently across different portions of the dataset and is not overly dependent on specific data points.

We cannot add an unlimited number of features due to the curse of dimensionality, as data becomes increasingly sparse in higher dimensions. This can lead to higher computational complexity without significant improvement in accuracy, and may even degrade model performance. For this, we can select only those features that have significant variance and contribute meaningfully to classification.