**Gaussian Discriminant Analysis for eCommerce, A Focus on Single-Variable Classification**

In diving deeper into Gaussian Discriminant Analysis (GDA), I'll start with a simpler case where we have just one variable, X. I'll get more mathematical here to illustrate this point. The Gaussian density function's mathematical form for class K when there’s a single X is given by a set of constants. The key part that depends on X appears in the exponential term. In this context, μₖ represents the mean for the observations in class K, and σₖ is the variance of that variable in class K. Initially, I’ll assume that the variance, σₖ, is the same across all classes. This is more than just a simplification—it's a crucial convenience that determines whether the discriminant function results in linear or quadratic forms.

By plugging the Gaussian density into Bayes' formula (as outlined earlier), I arrive at a rather complex expression. Essentially, I've inserted the form of the density into the numerator and included the summation across all classes in the denominator, making the expression appear quite intimidating. Fortunately, there are simplifications that I can apply. To classify an observation, I don’t need to calculate the exact probabilities but only need to determine which one is the largest. Hence, by taking the logarithm (a useful trick whenever exponentials are involved) and discarding terms that don’t depend on K, many elements cancel out. This reduces the complex expression to a simpler one that involves X (our single variable), the mean and variance of the distribution, and the prior probabilities. Importantly, in this scenario, this simplifies to a linear function of X, consisting of a single coefficient for X and a constant term composed of other factors.

This approach provides me with a discriminant function for each class. If there are only two classes, further simplification is possible. Let's consider a situation where the probabilities of class 1 and class 2 are both 0.5. In this scenario, the decision boundary occurs precisely at (μ₁ + μ₂) / 2. Referencing the earlier example, if the priors are equal and the two distributions are Gaussian, the decision boundary is at zero because the two means are equidistant from zero. Intuitively, this makes sense as the decision boundary represents the threshold where I switch from classifying an observation into one class versus the other. You can demonstrate this by using the expression for each class and determining when one is larger than the other—this isn't too hard to verify.

A bit of confusion arises here due to a squared term that seems to disappear. If I expand the squared term, there would indeed be an x² term. However, this x² term appears in both the numerator and each term in the denominator, and because there are no class-specific coefficients in front of this x² term, they cancel out in the ratio. This is a key point where careful mathematical handling simplifies the analysis.

The situation changes when I work with actual data rather than idealized population distributions. Instead of neatly drawn density functions, I rely on histograms to visualize data distributions. For the Gaussian rule, I need to estimate the means and the common standard deviation from the data. For example, if the true means are -1.5 and 1.5, with an average mean of 0 and equal class probabilities of 0.5, I must estimate these parameters from observed data and then apply them to the rule.

To do this, I estimate the priors simply as the number of observations in each class divided by the total number of observations. The means for each class are computed using the sample mean within each class. A more precise notation is required here, summing over all i such that yᵢ equals K (where yᵢ records the class label), which effectively isolates those Xᵢ values in class K to compute the mean. The pooled variance estimate, denoted by σ², is slightly more complex. Assuming equal variance across classes, I use a formula that sums the squared differences between each Xᵢ and its class mean, summed over all classes and divided by (n - K). Alternatively, another way to conceptualize this is by separately estimating the sample variance for each class and averaging them using a weighted approach. This weight corresponds to the proportion of observations in each class relative to the total number.

Having these estimates allows me to compute the decision boundary. Unlike in the idealized example where it was exactly at zero, with estimated values from actual data, it might shift slightly—though it could still be quite close to zero.

Next, I will further explore Gaussian Discriminant Analysis and its applications in statistical learning, especially in relation to eCommerce variables and classification problems.