Intuition:

The goal is to maximise the separation between the classes. In other words, we are interested in the features that best separate the class. For every feature, compute the mean and standard deviation for every class.

In order to maximise the separation between classes, their mean has to be as high as possible and their standard deviation should be as low as possible.

Means of classes along the dimension f should be High i.e., $\left(\mu\_{A, f}-\mu\_{B, f}\right)^{2}$ should be High.

Variance of classes along the dimension f should be Low i.e, $n\_A\*\sigma\_{A,f}^{2}+n\_B\*\sigma\_{B, f}^{2}$ should be Low.

where A and B are the classes, and f represents a feature. n\_A and n\_B are the number of data points in class A and class B, respectively. We have multiplied the n\_A and n\_B with variance to capture the scatter as there can be a varying number of data points in classes.

In Fischer Discrimination Analysis, we assume (first assumption)\*\* that classes have a unimodal distribution if one class has two peaks and is distributed at both extreme ends, while the other class is distributed somewhere in the middle. In such a case, it would be difficult to discriminate between the classes. (The \*\*second assumption\*\* is that if a class's distribution is not normal, we use logs or the necessary transformation to produce a normal distribution shape.

This approach is different from the Principal Component Analysis (PCA) approach in the way that it just picks the features, i.e., the features that best separate the classes, instead of projecting the features to lower dimensional space. Both approaches are linear transformation techniques and reduces the dimensions in the data.

\*\*While in PCA, we can get D projections, i.e., as many dimensions are there in the data, in the Fischer Discriminant Analysis, we can get up to C-1 projections, i.e., the number of classes C minus 1.\*\*