

Voici les raisons :

- la règle de classification $x^T \hat{\Sigma}^{-1} (\hat{\mu}_+ - \hat{\mu}_-)$ maximise la séparation entre les classes en se basant sur une distribution gaussienne multivariée.
- La matrice de covariance commune inversée $\hat{\Sigma}^{-1}$ simplifie le calcul de distance entre les classes. En effet, elle indique que la variabilité et la corrélation entre les variables sont supposées être les mêmes pour toutes les classes.
- $\hat{\mu}_+$, $\hat{\mu}_-$, $\hat{\Sigma}$ sont des estimateurs sans biais.
- La règle de décision est linéaire : ce nom vient du fait que il faut comparer la fonction discriminante linéaire à un seuil. Voici une explication
 - la partie gauche de l'inégalité, $x^T \hat{\Sigma}^{-1} (\hat{\mu}_+ - \hat{\mu}_-)$, est une expression linéaire. elle projette les données sur une ligne qui maximise la séparation entre les moyennes des classes projetées.
 - la partie droite, on dénote surtout les termes logarithmiques où m est le nombre d'observation de la classe -1 et n le nombre total d'observation. Ce qui ajuste le seuil en fonction des proportions a priori des classes. Cela veut dire que 'on reflète mieux la fréquence relative des classes dans les données d'entraînement. On évite aussi le biais envers la classe plus fréquente.