Assume each class is generated by a normal distribution. Sample covariance $\sigma_{xy}^2$ is
\[
\sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)/(n - 1).
\]

\[
\mu_1 = [1.464, 0.244] \quad \Sigma_1 = \begin{bmatrix} 0.0301 & 0.0057 \\ 0.0057 & 0.0115 \end{bmatrix}
\]

\[
\mu_2 = [4.260, 1.326] \quad \Sigma_2 = \begin{bmatrix} 0.2208 & 0.0731 \\ 0.0731 & 0.0391 \end{bmatrix}
\]

\[
\mu_3 = [5.552, 2.026] \quad \Sigma_3 = \begin{bmatrix} 0.3046 & 0.0488 \\ 0.0488 & 0.0754 \end{bmatrix}
\]
Discriminant Functions

Discriminant functions $d_i(x)$ for classes $\omega_i$ are designed so that the decision for $x$ is class:

$$\arg \max_i P(\omega_i \mid x)$$

The decision boundary between two classes $\omega_i$ and $\omega_j$ is where $d_i(x) = d_j(x)$.

Typically, $d_i(x) \propto \ln P(x \mid \omega_i) + \ln P(\omega_i)$

For normal distributions, $d_i(x)$ is equal to:

$$-rac{1}{2} \ln |\Sigma_i| - \frac{1}{2}(x - \mu_i)\Sigma_i^{-1}(x - \mu_i)^T + \ln P(\omega_i)$$
Plot of decision boundaries assuming equal $P(\omega_i)$

Decision boundaries assuming $L_{i2} = 100$
Linear Discriminant Functions

If we assume equal covariance matrices, then the boundaries are linear instead of quadratic.

\[ d_i(x) = -\frac{1}{2}(x - \mu_i)\Sigma^{-1}(x - \mu_i)^T + \ln P(\omega_i) \]

\((x - \mu_i)\Sigma^{-1}(x - \mu_i)^T\) is the Mahalanobis distance. By dropping common terms, \(d_i(x)\) simplifies to:

\[ d_i(x) = \mu_i\Sigma^{-1}x_i^T - \frac{1}{2}\mu_i\Sigma^{-1}\mu_i^T + \ln P(\omega_i) \]

Decision boundaries for equal covariance

\[ \Sigma = \begin{bmatrix} 0.1852 & 0.0425 \\ 0.0425 & 0.0420 \end{bmatrix} \]
Comments on Statistical Approach

Advantages: efficient, rarely overfit, often very good performance, easy to interpret

Disadvantage: Assumptions about prob. dists. might be wrong or too simple, leading to underfitting.

NNs and SVMs provide models with more flexibility. It will be important to ensure efficiency and avoid overfitting.