ML Problem Solving

Multivariate Gaussian Distribution

1. Consider the following dataset, where each row is an observation. First and second columns are input features and the true output is the last column.

$D = [[3, 3, 1], [1, 1, 1], [-1, 0, 1], [2, 2, 0], [-2, 2, 0], [-2, -2, 0], [0, -2, 0]]$

Draw the points in a 2-D coordinate system.

2. Calculate the μ mean vector and the Σ covariance matrix from the data points that have class = 1.

3. The Mahalanobis distance d_M is the multivariate generalization of the distance (in standard deviations) a point is from the mean of a distribution. It is given by $d_M = \sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)}$. Assume $X \sim \mathcal{N}(\mu, \Sigma)$ with estimated μ and Σ. Find 4 points that have $d_M = 1.0$ and draw them in your coordinate system. What can we say about the points with the same Mahalanobis distance?

4. Find the probability density $Pd(X|class = 1)$ associated with the 4 points used in the previous question.

$Pd(X|class) = \frac{1}{(2\pi)^{F/2}} \frac{1}{|\Sigma|^{1/2}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$

5. What is the most probable point in the input space, considering the same probability distribution $Pd(X|class = 1)$?