2/10/23
Day 9

The Multivariate Gaussian distribution

- Each predictor is μ, σ distributed, w/ some correlation.

We write this as $X \sim N(\mu, \Sigma)$, where Σ is the covariance matrix.

$$f(x) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

In the case of $p > 1$ predictors, LDA assumes that observations in the k-th class are drawn from a ~~normal~~ $N(\mu_k, \Sigma)$ distribution, where μ_k is class specific and Σ is a covariance matrix common to all classes.

One can show that here, the discriminant function becomes

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k$$

μ_ℓ and π_ℓ $(\ell = 1, \dots, k)$ are estimated as before, as is Σ

- The higher the ratio P/n, the more of a problem is overfitting.

A binary classifier can make two types of error
1) Assign a true class 1 to class 2
2) Assign a true class 2 to class 1.

Confusion matrix

$$\begin{array}{ll|cc} & & \text{True} & \text{class} \\ & & 1 & 2 \\ \hline \text{Predicted} & 1 & 9644 & 252 \\ \text{class} & 2 & 23 & 81 \end{array}$$

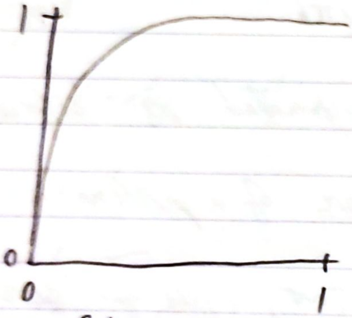
- **Sensitivity:** the percentage of true defaulters who are identified
- **Specificity:** the percentage of non-defaulters that are identified.

- The sensitivity of this model is only 24.3%
- The specificity is 99.8%.
- The credit card company would rather have high sensitivity than specificity; we can modify LDA for this purpose.
  ↳ we might label any customer with a posterior probability of default above 20% to the default class, i.e. if
  $\mathbb{P}(\text{default} = \text{Yes} \mid X = x) > 0.2$

Choosing the threshold depends on domain knowledge.

- The ROC curve (receiver operating characteristics) simultaneously displays the two types of errors for all possible thresholds.

True positive rate



False positive rate

- The overall performance of a classifier is given by the area under the (ROC) curve, abbreviated as AUC

The closer AUC is to 1, the better, ~~you want~~ because that means the ROC curve hugs the upper left, meaning the true positive rate is very high in exchange for only a small increase in false positive rate.

True positive rate = sensitivity
False positive rate = 1 - sensitivity

False Positive rate = type I error = 1 - specificity

True positive rate = 1 - type II error = power = sensitivity = recall

Positive predictive value = precision = 1 - false discovery proportion

Negative predictive value