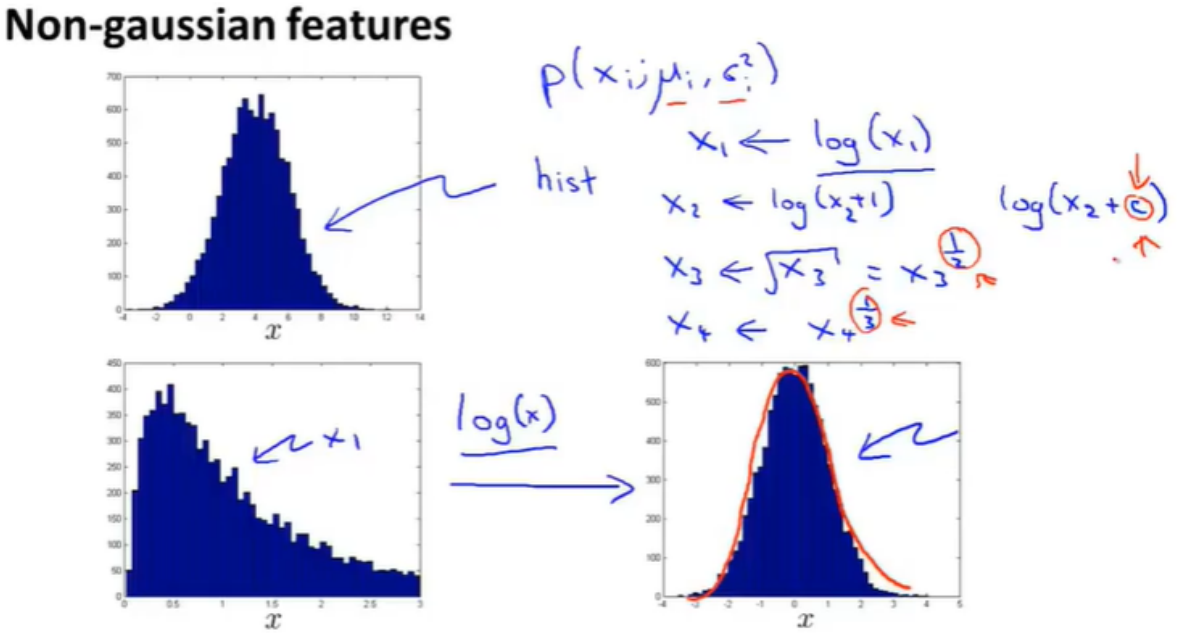
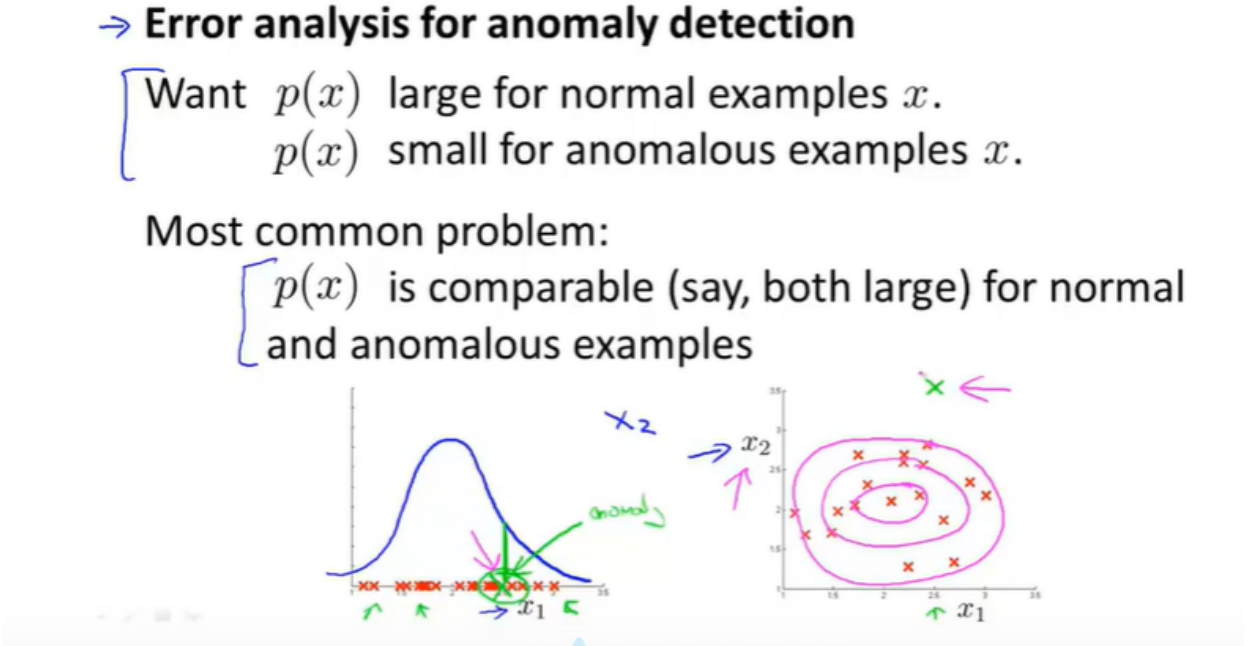
第一步：先将数据进行一些处理，让它看起来像是高斯分布。



第二步：进行**误差分析**，先训练出一个模型，然后在一组交叉验证集上找出出错的样本，尝试一些其他特征，看看能否纠正这些错误。



逐个加入变量，直到降低了异常样本的p(x)值.

选择特征明显的特征值

- **Monitoring computers in a data center**
- Choose features that might take on unusually large or small values in the event of an anomaly.
  - x_1 = memory use of computer
  - x_2 = number of disk accesses/sec
  - x_3 = CPU load
  - x_4 = network traffic

$$x_5 = \frac{\text{CPU load}}{\text{network traffic}}$$

$$x_6 = \frac{(\text{CPU load})^2}{\text{network traffic}}$$