

HW1 09/08/18

Problem 1.

1A: 用高斯分布来拟合 feature 的分布. 1. 高斯分布模型时考虑 mean, var (1, 6)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

取 log.

$$\begin{aligned}\log f(x) &= \log(\sigma\sqrt{2\pi})^{-1} + \log e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= -\log\sqrt{2\pi}\sigma + \left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\ &= -\log\sqrt{2\pi} - \log\sigma - \frac{(x-\mu)^2}{2\sigma^2} \\ &= -\frac{(x-\mu)^2}{2\sigma^2} - \log\sigma - \log\sqrt{2\pi}\end{aligned}$$

形状不一致的向量

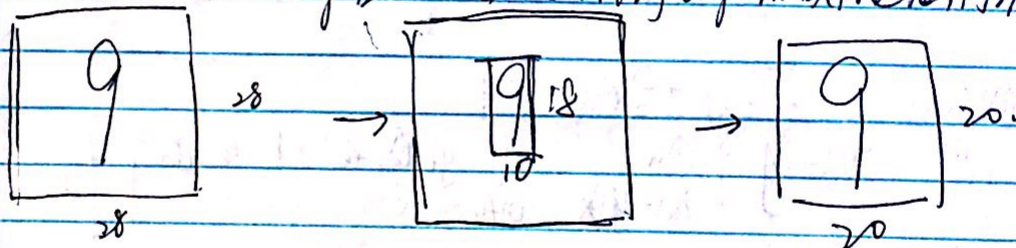
1B: 将 dataset 中 null 值改成 NA

1D: 直接调用 `sum light` 来完成

Problem 2.

2A - 用两个 model 来拟合. Gaussian Distribution / Bernoulli Distribution

- Stretched Bounding Box 类似先标注 digit 的实际位置, 再缩放至 28×28



- 画 mean image: 是将 test 后的分类结果, 同个 digit 取平均. 主要是用来可视化 test 结果后.

2B Random Forest 测试没啥特别的.

