50.039 Theory and Practice of Deep Learning
Theory Homework 1

Joel Huang 1002530

February 3, 2019

Drawing data from a distribution of (x, y)

Problem definition

- We need to draw pairs of data x and binary labels $y = 0$ or $y = 1$.
- There are two Gaussians with cluster index $c(x) = 1$ and $c(x) = 2$, each with a different probability of getting class $y = 0$ and $y = 1$.
- x follows the Gaussian distributions of the two clusters.

Formulation

- x and y are independent: $P(x, y) = P(x)\,P(y)$.
- y is only dependent on cluster index $c(x)$. So $P(y \mid x, c(x)) = P(y \mid c(x))$.
- $P(x)$ is the product of the distribution of the data points given the cluster, and the probability of choosing the cluster:

$$\begin{aligned} P(x) =& f(x \mid c(x))\,P(c(x)) \\ =& f(x \mid c(x) = 1)\,P(c(x) = 1) \\ &+ f(x \mid c(x) = 2)\,P(c(x) = 2) \end{aligned}$$

- $P(y = 0)$ is the product of the probability of generating label $y = 0$ given the cluster, and the probability of choosing the cluster:

$$\begin{aligned} P(y = 0) =& P(y = 0 \mid c(x) = 1)\,P(c(x) = 1) \\ &+ P(y = 0 \mid c(x) = 2)\,P(c(x) = 2) \end{aligned}$$

- $P(y = 1)$ can be expressed in terms of $P(y = 0)$ because the labels are binary. Using the relationship $P(y = 1 \mid c(x)) = 1 - P(y = 0 \mid c(x))$:

$$\begin{aligned} P(y = 1) =& (1 - P(y = 0 \mid c(x) = 1))\,P(c(x) = 1) \\ &+ (1 - P(y = 0 \mid c(x) = 2))\,P(c(x) = 2) \end{aligned}$$

Expressing $P(x, y)$

$$\begin{aligned} P(x, y) &= P(x)P(y) \\ &= \begin{cases} P(x)P(y = 0), & \text{for } y = 0 \\ P(x)P(y = 1), & \text{for } y = 1 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

For $y = 0$,

$$\begin{aligned} P(x, y = 0) =& P(x)P(y = 0) \\ =& f(x \mid c(x))\,P(c(x)) \\ &\cdot [(P(y = 0 \mid x, c(x) = 1)\,P(c(x) = 1) \\ &+ P(y = 0 \mid x, c(x) = 2)\,P(c(x) = 2)] \\ =& [f(x \mid c(x) = 1)\,P(c(x) = 1) \\ &+ f(x \mid c(x) = 2)\,P(c(x) = 2)] \\ &\cdot [0.2 \cdot P(c(x) = 1) + 0.7 \cdot P(c(x) = 2)] \end{aligned}$$

Similarly for $y = 1$,

$$\begin{aligned} P(x, y = 1) =& P(x)P(y = 1) \\ =& f(x \mid c(x))\,P(c(x)) \\ &\cdot [(1 - (P(y = 0 \mid x, c(x) = 1))\,P(c(x) = 1) \\ &+ (1 - P(y = 0 \mid x, c(x) = 2))\,P(c(x) = 2)] \\ =& [f(x \mid c(x) = 1)\,P(c(x) = 1) \\ &+ f(x \mid c(x) = 2)\,P(c(x) = 2)] \\ &\cdot [0.8 \cdot P(c(x) = 1) + 0.3 \cdot P(c(x) = 2)] \end{aligned}$$