50.039 Theory and Practice of Deep Learning
Theory Homework 1

Joel Huang 1002530

February 4, 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

- $P(x, y) = P(y, x, c(x) = 1) + P(y, x, c(x) = 2)$.
- y is only dependent on cluster index $c(x)$. So $P(y \,|\, x, c(x)) = P(y \,|\, c(x))$.
- $P(x, c(x))$ can be decomposed into the product of the distribution of the data points given the cluster, and the probability of choosing the cluster:

$$P(x, c(x) = 1) = f(x \,|\, c(x) = 1)\, P(c(x) = 1)$$

$$P(x, c(x) = 2) = f(x \,|\, c(x) = 2)\, P(c(x) = 2)$$

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

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

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

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

Expressing $P(x, y)$

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

For $y = 0$,

$$\begin{aligned} P(x, y = 0) =& P(y = 0 \,|\, x, c(x) = 1)\, P(x, c(x) = 1) \\ &+ P(y = 0 \,|\, x, c(x) = 2)\, P(x, c(x) = 2) \\ =& 0.2 \cdot f(x \,|\, c(x) = 1)\, P(c(x) = 1) \\ &+ 0.7 \cdot f(x \,|\, c(x) = 2)\, P(c(x) = 2) \\ =& 0.2 \cdot 0.5 \cdot f(x \,|\, c(x) = 1) \\ &+ 0.7 \cdot 0.5 \cdot f(x \,|\, c(x) = 2) \end{aligned}$$

Similarly for $y = 1$,

$$\begin{aligned} P(x, y = 1) =& P(y = 1 \,|\, x, c(x) = 1)\, P(x, c(x) = 1) \\ &+ P(y = 1 \,|\, x, c(x) = 2)\, P(x, c(x) = 2) \\ =& (1 - P(y = 0 \,|\, x, c(x) = 1))\, P(x, c(x) = 1) \\ &+ (1 - P(y = 0 \,|\, x, c(x) = 2))\, P(x, c(x) = 2) \\ =& 0.8 \cdot f(x \,|\, c(x) = 1)\, P(c(x) = 1) \\ &+ 0.3 \cdot f(x \,|\, c(x) = 2)\, P(c(x) = 2) \\ =& 0.8 \cdot 0.5 \cdot f(x \,|\, c(x) = 1) \\ &+ 0.3 \cdot 0.5 \cdot f(x \,|\, c(x) = 2) \end{aligned}$$