

1. Theory Question: In the lecture we encountered the plot reproduced here, showing a structured classification problem. It was pointed out that the most salient feature of this problem is that each class seems to have its own generative distribution.

i.e. the blue points have a different distribution than the green ones, (although both distributions happen to overlap).

Such situations indicate an anti-causal relationship.
("the label causes the inputs")

and it might seem that discriminative modeling paradigm adopted in the lecture is not appropriate.

The goal of this week's theory is to realise that the situation is a bit more subtle than that.

For the purpose of this exercise, we will assume that there are two classes C_1 and C_2 , defining probability distributions

$p(x|C_1)$, $p(x|C_2)$ over the inputs, and classes are drawn with the probability.

$$p(C) = [p(C_1), p(C_2) = 1 - p(C_1)]$$

(a) Given a new input x , how would you compute the posterior $P(C_1|x)$? Show that it can be written as a logistic function.

Ans: $P(C_1|x) = \frac{1}{1 + e^{-\alpha(x)}}$ where $\alpha(x) = \ln \frac{P(C_1|x)}{P(C_2|x)}$

Consider the RHS

$$\Rightarrow \frac{1}{1 + e^{-\alpha(x)}} = \frac{1}{1 + e^{-\ln \frac{P(C_1|x)}{P(C_2|x)}}} = \frac{1}{1 + \frac{P(C_2|x)}{P(C_1|x)}} = \frac{P(C_1|x)}{P(C_1|x) + P(C_2|x)}$$

using Bayes rule

$$\Rightarrow \frac{\frac{P(C_1|x)}{P(x|C_1)p(C_1)}}{\frac{P(x|C_1)p(C_1)}{P(x)} + \frac{P(x|C_2)p(C_2)}{P(x)}}$$

$$\Rightarrow P(C_1|x) \left[\frac{P(x)}{P(x|C_1)p(C_1) + P(x|C_2)p(C_2)} \right]$$

From Law of total probability $P(x) = P(x|C_1)p(C_1) + P(x|C_2)p(C_2)$

$$\Rightarrow P(C_1|x) \left[\frac{P(x)}{P(x)} \right] = P(C_1|x)$$

1(b). This observation suggests that when we perform logistic regression to learn the function $a(x)$, we actually indirectly learn the class distributions $p(x|C_1)$ and $p(x|C_2)$. It is interesting to consider how different regression models for $a(x)$ relate to different assumptions about class distributions. Assume that both classes are drawn from the same exponential family, with different parameters.

$$p(x|C_k) = h(x) \exp(\phi(x)^T \omega_k - \log Z(\omega_k))$$

Show that this implies a linear model for $a(x)$

$$a(x) = \phi(x)^T \vec{\theta} + \theta_0$$

What are the parameters $\vec{\theta}$ and θ_0 of this model in terms of the parameters ω of the class distributions?

Answer:

① Let $p(x|C_1) = h(x) \exp(\phi(x)^T \omega_1 - \log Z(\omega_1))$
 and $p(x|C_2) = h(x) \exp(\phi(x)^T \omega_2 - \log Z(\omega_2))$

then we have $\frac{p(C_1|x)}{p(C_2|x)} = \frac{\frac{p(x|C_1) p(C_1)}{p(x)}}{\frac{p(x|C_2) p(C_2)}{p(x)}} = \frac{p(x|C_1)}{p(x|C_2)} \frac{p(C_1)}{p(C_2)}$

$$\Rightarrow \frac{h(x) \exp(\phi^T(x) \omega_1 - \log Z(\omega_1))}{h(x) \exp(\phi^T(x) \omega_2 - \log Z(\omega_2))} \frac{p(C_1)}{p(C_2)}$$

$$\Rightarrow \exp[\phi^T(x)(\omega_1 - \omega_2) - \log Z(\omega_1) + \log Z(\omega_2)] \frac{p(C_1)}{p(C_2)}$$

$$\Rightarrow \exp[\phi^T(x)(\omega_1 - \omega_2) - \ln \frac{Z(\omega_1)}{Z(\omega_2)} + \ln \frac{p(C_1)}{p(C_2)}]$$

We have $a(x) = \ln \frac{p(c_1|x)}{p(c_2|x)}$

Thus

$$a(x) = \ln \left[\exp \left\{ \phi^T(x) (\omega_1 - \omega_2) - \ln \frac{Z(\omega_1)}{Z(\omega_2)} + \ln \frac{p(c_1)}{p(c_2)} \right\} \right]$$

$$a(x) = \phi^T(x) (\omega_1 - \omega_2) + \left[\ln \frac{p(c_1)}{p(c_2)} - \ln \frac{Z(\omega_1)}{Z(\omega_2)} \right]$$

thus $\vec{\theta} = \omega_1 - \omega_2$

$$\theta_0 = \ln \frac{p(c_1)}{p(c_2)} - \ln \frac{Z(\omega_1)}{Z(\omega_2)}$$

and $a(x) = \vec{\phi}^T(x) \vec{\theta} + \theta_0$ is a linear model

- (c) Does this mean we can turn any discriminative model (predicting classlabels from inputs, $p(C_k | \mathbf{x})$) into a generative one (predicting inputs from class labels, $p(\mathbf{x} | C_k)$) by "going backward" through the model? Unfortunately, the answer is *no*, because continuous distributions for inputs are more complex than binary distributions. To make this clear, consider a special case: Assume that the class distributions are both univariate Gaussians with different means and variances: $p(\mathbf{x} | C_k) = \mathcal{N}(\mathbf{x}; \mu_k, \sigma_k^2)$. From your answer above, construct the explicit form of θ, θ_0 (that's two real numbers) as a function of $\mu_1, \mu_2, \sigma_1, \sigma_2, p(C_1)$. Assume someone has performed logistic regression and given you the parameters θ, θ_0 . Can you recover the parameters $\mu_1, \mu_2, \sigma_1, \sigma_2, p(C_1)$? Could you do so if we set $p(C_1) = p(C_2) = \frac{1}{2}$ and $\sigma = 1$?

Answer: $p(C_k | \mathbf{x})$ into $p(\mathbf{x} | C_k)$

inputs \rightarrow class labels. class labels \rightarrow inputs

* continuous distributions for inputs are more complex than binary distributions.

We assume that class distributions are both univariate Gaussians with different mean and variances.

$$p(x|C_1) = \mathcal{N}(x; \mu_1, \sigma_1^2) \quad p(x|C_2) = \mathcal{N}(x; \mu_2, \sigma_2^2)$$

we know

$$\begin{aligned} p(x|\omega) &= \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right] \\ &= \exp \left[-\frac{x^2 - 2\mu x + \mu^2}{2\sigma^2} - \log \sqrt{2\pi\sigma^2} \right] \\ &= \exp \left[[x \quad -\frac{1}{2}x^2] \begin{bmatrix} \mu/\sigma^2 \\ 1/\sigma^2 \end{bmatrix} - \left(\frac{\mu^2}{2\sigma^2} + \log \sqrt{2\pi\sigma^2} \right) \right] \end{aligned}$$

$$\text{if } \omega_1 = \mu/\sigma^2 \quad \omega_2 = 1/\sigma^2$$

$$= \exp \left[[x \quad -\frac{1}{2}x^2] \begin{bmatrix} \omega_1 \\ \omega_2 \end{bmatrix} - \underbrace{\frac{1}{2} \left(\frac{\omega_1^2}{\omega_2} - \log \omega_2 + \log (2\pi) \right)}_{\log Z(\omega)} \right]$$

$$P(x|C_1) = \exp \left[[x - \mu_{x^2}] \begin{bmatrix} \mu_1/\sigma_1^2 \\ 1/\sigma_1^2 \end{bmatrix} - \frac{1}{2} \left(\frac{\mu_1^2}{\sigma_1^2} - \log 2\pi\sigma_1^2 \right) \right]$$

$$P(x|C_2) = \exp \left[[x - \mu_{x^2}] \begin{bmatrix} \mu_2/\sigma_2^2 \\ 1/\sigma_2^2 \end{bmatrix} - \frac{1}{2} \left(\frac{\mu_2^2}{\sigma_2^2} - \log 2\pi\sigma_2^2 \right) \right]$$

$$\frac{P(C_1|x)}{P(C_2|x)} = \frac{\overbrace{P(C_1)}^{\text{constant}}}{\overbrace{P(C_2)}^{\text{constant}}} \frac{\exp \left[[x - \mu_{x^2}] \begin{bmatrix} \mu_1/\sigma_1^2 \\ 1/\sigma_1^2 \end{bmatrix} - \frac{1}{2} \left(\frac{\mu_1^2}{\sigma_1^2} - \log 2\pi\sigma_1^2 \right) \right]}{\exp \left[[x - \mu_{x^2}] \begin{bmatrix} \mu_2/\sigma_2^2 \\ 1/\sigma_2^2 \end{bmatrix} - \frac{1}{2} \left(\frac{\mu_2^2}{\sigma_2^2} - \log 2\pi\sigma_2^2 \right) \right]}$$

$$= \frac{P(C_1)}{P(C_2)} \exp \left[[x - \mu_{x^2}] \begin{bmatrix} \mu_1/\sigma_1^2 - \mu_2/\sigma_2^2 \\ 1/\sigma_1^2 - 1/\sigma_2^2 \end{bmatrix} - \frac{1}{2} \left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2} - \log \frac{\sigma_1^2}{\sigma_2^2} + \log 2\pi\sigma_2^2 \right) \right]$$

$$= \exp \left[[x - \mu_{x^2}] \begin{bmatrix} \mu_1/\sigma_1^2 - \mu_2/\sigma_2^2 \\ 1/\sigma_1^2 - 1/\sigma_2^2 \end{bmatrix} - \frac{1}{2} \left(\frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2} - \log \frac{\sigma_1^2}{\sigma_2^2} + \log 2\pi\sigma_2^2 \right) + \ln \left[\frac{P(C_1)}{P(C_2)} \right] \right]$$

Consider ratio of class prior probabilities.

$$\ln \frac{P(C_1)}{P(C_2)} \quad \text{and variances} \quad -\frac{1}{2} \log \frac{\sigma_1^2}{\sigma_2^2}$$

for all values where these terms coincide
 $\alpha(x)$ will be the same thus $P(C_1), P(C_2)$
 σ_1, σ_2 will not be recognizable.

if we have $\sigma_1 = \sigma_2 = 1$ and $P(c_1) = P(c_2) = 1/2$

$$a(x) = \begin{bmatrix} x & -1/x^2 \end{bmatrix} \begin{bmatrix} \mu_1 - \mu_2 \\ 0 \end{bmatrix} - \frac{1}{2} [\mu_1^2 - \mu_2^2]$$

we have been given that θ, θ_0 are the
the result of performing logistic regression then
we have-

$$\begin{bmatrix} \theta \\ \theta_0 \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ 0 \end{bmatrix} \quad \text{and} \quad \theta_0 = -\frac{1}{2} \mu_1^2 + \mu_2^2$$

$$\theta_1 = \mu_1 - \mu_2$$

$$\theta_0 = \frac{1}{2} (\mu_1 + \mu_2) (\theta_1)$$

$$\frac{\theta_0}{\theta_2} = \frac{1}{2} (\mu_1 + \mu_2)$$

Thus μ_1 and μ_2 are recoverable in
this scenario