# Lecture 12: Logistic Regression

LDA: $\delta_c(\underset{\sim}{x}) = \mathbb{P}(Y=c \mid \underset{\sim}{X} = \underset{\sim}{x})$

$$\propto \underbrace{\mathbb{P}(\underset{\sim}{X}=\underset{\sim}{x} \mid Y=c)}_{N(\mu_c, \Sigma)} \underbrace{\mathbb{P}(Y=c)}_{\pi_c}$$

## Logistic Reg.:

$$\delta_c(\underset{\sim}{x}) = \mathbb{P}(Y=c \mid \underset{\sim}{X} = \underset{\sim}{x})$$

model this directly

---

## Binary Logistic Regression ($K=2$)

We have $Y = 0$ or $1$

$\delta_0(\underset{\sim}{x}) = \mathbb{P}(Y=0 \mid \underset{\sim}{X}=\underset{\sim}{x}) = 1 - \mathbb{P}(Y=1 \mid \underset{\sim}{X}=\underset{\sim}{x}) = 1-\delta_1(\underset{\sim}{x})$

$\delta_1(\underset{\sim}{x}) = \mathbb{P}(Y=1 \mid \underset{\sim}{X}=\underset{\sim}{x}) = \dots = 1-\delta_0(\underset{\sim}{x})$

So I really only need one,

predict class 1 when $\delta_1(\underset{\sim}{x}) > \delta_0(\underset{\sim}{x})$

i.e. $\delta_1(\underset{\sim}{x}) > 1-\delta_1(\underset{\sim}{x})$

i.e. $\delta_1(\underset{\sim}{x}) > 1/2$

Call $p(\underset{\sim}{x}) = \delta_1(\underset{\sim}{x}) = \mathbb{P}(Y=1 \mid \underset{\sim}{X}=\underset{\sim}{x})$

Given $\underset{\sim}{X} = \underset{\sim}{x}$, $Y = 0$ or $1$.

i.e. $$Y \mid \underset{\sim}{X} = \underset{\sim}{x} \sim \text{Bern}(p(\underset{\sim}{x}))$$

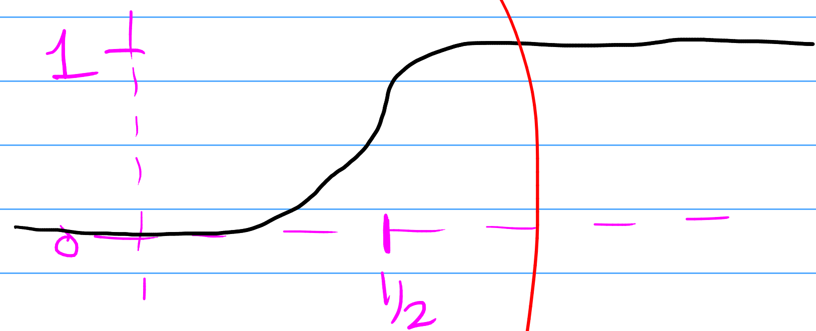
<u>Game</u>: model $p(\underset{\sim}{x})$ in a reasonable way.

$p(\underset{\sim}{x}) = \hat{\beta}^T \underset{\sim}{x}$ i.e. linear in $\underset{\sim}{x}$


Logistic regression says let

$$p(\underset{\sim}{x}) = \text{logistic}(\hat{\beta}^T \underset{\sim}{x})$$

$$\text{logistic}(x) = \frac{1}{1+e^{-x}}$$




$$\text{logit}(x) = \text{logistic}^{-1}(x)$$
$$= \log\left(\frac{x}{1-x}\right)$$

$$p(\underset{\sim}{x}) = \text{logistic}(\hat{\beta}^T \underset{\sim}{x})$$

$$= \frac{1}{1 + \exp(-\hat{\beta}^T \underset{\sim}{x})}$$

$\underset{\sim}{x} = (1, x_1, x_2, \ldots, x_p)$

$$= \frac{1}{1 + \exp(-(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_p x_p))}$$

Notice:

log-odds

$$\text{logit}(p(\underset{\sim}{x})) = \hat{\beta}^T \underset{\sim}{x}$$

$$\Rightarrow \log\left(\underbrace{\frac{p(\underset{\sim}{x})}{1 - p(\underset{\sim}{x})}}_{\text{odds}}\right) = \hat{\beta}^T \underset{\sim}{x}$$

---

How do we get $\hat{\beta}$?

$p_\beta(\underset{\sim}{x}_n) = \frac{1}{1 + e^{-\beta^T x_n}}$

$$Y_n \mid \underset{\sim}{X}_n = \underset{\sim}{x}_n \overset{\text{indep}}{\sim} \text{Bern}(p_\beta(\underset{\sim}{x}_n))$$

Get $\hat{\beta}$ as a MLE (maximum likelihood estimate)

$$L(\beta) = \mathbb{P}(Y_1, Y_2, Y_3, \ldots, Y_N \mid \underset{\sim}{X}_1 = \underset{\sim}{x}_1, \ldots, \underset{\sim}{X}_n = \underset{\sim}{x}_n)$$

$$= \prod_{n=1}^{N} \mathbb{P}(Y_n = y_n \mid \underset{\sim}{X}_n = \underset{\sim}{x}_n)$$

$Z \sim \text{Bern}(p)$

$$f(z) = p^z (1-p)^{1-z}$$

$$= \prod_{n=1}^{N} p_\beta(\underset{\sim}{x}_n)^{y_n} (1 - p_\beta(\underset{\sim}{x}_n))^{1-y_n}$$

$$L(\beta) = \prod_{n=1}^{N} \left( \frac{1}{1+e^{-\beta^T x_n}} \right)^{y_n} \left( 1 - \frac{1}{1+e^{-\beta^T x_n}} \right)^{1-y_n}$$

**MLE:** $\hat{\beta} = \underset{\beta}{\arg\max}\ L(\beta)$

$\rightarrow$ no analytical soln $\longrightarrow$ use optimization approach / software

---

## Multinomial Logistic Regression (multi-class logistic $K>2$)

$$Y_n \mid \underset{\sim}{X}_n = \underset{\sim}{x}_n \overset{indep}{\sim} \text{MultiNomial}(p_{1\beta_1}(\underset{\sim}{x}_n), p_{2\beta_2}(\underset{\sim}{x}_n), \ldots, p_{K-1\,\beta_{K-1}}(\underset{\sim}{x}_n))$$

For $k = 1, \ldots, K-1$

$$\delta_k(\underset{\sim}{x}) = p_k(\underset{\sim}{x}) = \mathbb{P}(Y = k \mid \underset{\sim}{X} = \underset{\sim}{x})$$

model as multivariate logistic fn

$$= \frac{\exp(\beta_k^T \underset{\sim}{x})}{1 + \sum_{l=1}^{K-1} \exp(\beta_l^T \underset{\sim}{x})}$$

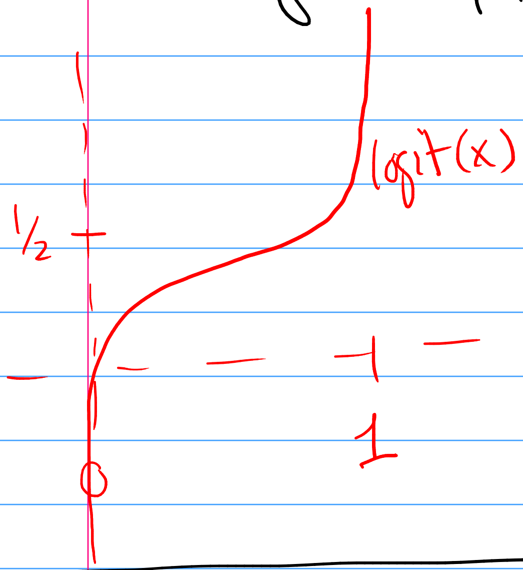
We similarly estimate $\beta_k$ as the MLEs.

---

Back to Binary Logistic Regr ($K=2$)

$$\text{logit}(\delta_1(\underset{\sim}{x})) = \hat{\beta}^T \underset{\sim}{x}$$

linear classifier.




---

## LDA v. Logistic Regression

| LDA | Logistic Regression |
| --- | --- |
| (1) models $P(X\|Y)$ and $P(Y)$ using a normality assumption about $X\|Y$ | (1) models $Y\|X$ directly makes no assumption about X (more general) |
| (2) normality makes fitting much easier | (2) relatively harder to fit |

(3) both linear