ECE595 / STAT598: Machine Learning I
Lecture 15.2: Logistic Regression 2 - Connection with Bayes

Spring 2020

Stanley Chan

School of Electrical and Computer Engineering
Purdue University

Outline

Discriminative Approaches

- Lecture 14 Logistic Regression 1
- Lecture 15 Logistic Regression 2

This lecture: Logistic Regression 2

- Gradient Descent
  - Convexity
  - Gradient
  - Regularization
- Connection with Bayes
  - Derivation
  - Interpretation
- Comparison with Linear Regression
  - Is logistic regression better than linear?
  - Case studies

Connection with Bayes

- The likelihood is

$$p(\boldsymbol{x}|i) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i)\right\}$$

- The prior is $p_Y(i) = \pi_i$.
- The posterior is

$$\begin{aligned} p(1|\boldsymbol{x}) &= \frac{p(\boldsymbol{x}|1)p_Y(1)}{p(\boldsymbol{x}|1)p_Y(1) + p(\boldsymbol{x}|0)p_Y(0)} \\ &= \frac{1}{1 + \frac{p(\boldsymbol{x}|0)p_Y(0)}{p(\boldsymbol{x}|1)p_Y(1)}} = \frac{1}{1 + \exp\left\{-\log\left(\frac{p(\boldsymbol{x}|1)p_Y(1)}{p(\boldsymbol{x}|0)p_Y(0)}\right)\right\}} \\ &= \frac{1}{1 + \exp\left\{-\log\left(\frac{\pi_1}{\pi_0}\right) - \log\left(\frac{p(\boldsymbol{x}|1)}{p(\boldsymbol{x}|0)}\right)\right\}}. \end{aligned}$$

Connection with Bayes

- We can show that the last term is

$$
\begin{aligned}
&\log\left(\frac{p(\boldsymbol{x}|1)}{p(\boldsymbol{x}|0)}\right) \\
&= \log\left(\frac{\frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)\right\}}{\frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\right\}}\right) \\
&= -\frac{1}{2}\left[(\boldsymbol{x}-\boldsymbol{\mu}_1)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)-(\boldsymbol{x}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_0)\right] \\
&= (\boldsymbol{\mu}_1-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}^{-1}\boldsymbol{x}-\frac{1}{2}\left(\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1-\boldsymbol{\mu}_0^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_0\right).
\end{aligned}
$$

- Let us define

$$
\begin{aligned}
\boldsymbol{w} &= \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_0) \\
w_0 &= -\frac{1}{2}\left(\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1-\boldsymbol{\mu}_0^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_0\right)+\log\left(\frac{\pi_1}{\pi_0}\right)
\end{aligned}
$$

Connection with Bayes

- Then,

$$\log\left(\frac{p(\boldsymbol{x}|1)}{p(\boldsymbol{x}|0)}\right) = (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \frac{1}{2}\left(\boldsymbol{\mu}_1^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_0\right)$$
$$= \boldsymbol{w}^T\boldsymbol{x} + w_0 - \log \pi_1/\pi_0$$

- Therefore,

$$p(1|\boldsymbol{x}) = \frac{1}{1+\exp\left\{-\log\left(\frac{\pi_1}{\pi_0}\right) - \log\left(\frac{p(\boldsymbol{x}|1)}{p(\boldsymbol{x}|0)}\right)\right\}}$$
$$= \frac{1}{1+\exp\{-(\boldsymbol{w}^T\boldsymbol{x} + w_0)\}}$$
$$= h_{\boldsymbol{\theta}}(\boldsymbol{x})$$

Connection with Bayes

- The hypothesis function is the posterior distribution

$$
\begin{aligned}
p_{Y|\boldsymbol{X}}(1|\boldsymbol{x}) &= \frac{1}{1+\exp\{-(\boldsymbol{w}^T\boldsymbol{x}+w_0)\}} = h_{\boldsymbol{\theta}}(\boldsymbol{x}) \\
p_{Y|\boldsymbol{X}}(0|\boldsymbol{x}) &= \frac{\exp\{-(\boldsymbol{w}^T\boldsymbol{x}+w_0)\}}{1+\exp\{-(\boldsymbol{w}^T\boldsymbol{x}+w_0)\}} = 1-h_{\boldsymbol{\theta}}(\boldsymbol{x}),
\end{aligned} \tag{1}
$$

- So logistic regression offers probabilistic reasoning which linear regression does not
- Not true when the covariances are different
- Remark: If the covariances are different, the Bayes returns a quadratic classifier