

Discriminative Model

Logistic/SVM/Tree

(no assumptions on $p(x)$)



$$\begin{aligned} P(x, y) &= P(Y=y \mid X=x) P(X=x) \\ &= P(X=x \mid Y=y) P(Y=y) \end{aligned}$$

Generative Model

Discriminant Analysis

LDA/QDA/NB



Logistic Regression (Binary)

range = (0, 1)

$$\eta(x) = P(Y = 1|X = x)$$
$$1 - \eta(x) = P(Y = 0|X = x)$$

range = (-inf, inf)

$$\log \frac{\eta(x)}{1 - \eta(x)} = x^t \beta$$

Here beta contains
the intercept

At each data point, we have $(x_i, y_i, \eta(x_i))$

How to estimate eta(x), or equivalently beta?

Choice of Loss function $L(y, \eta(x))$:

$\rightarrow (y - \eta(x))^2$, **log-likelihood**?

Too flat to train

MLE for Binary Logistic Regression

$$\log \frac{\eta(x)}{1 - \eta(x)} = x^t \beta$$

$$P(Y = 1 | X = x) = \eta(x) = \frac{\exp(x^t \beta)}{1 + \exp(x^t \beta)}$$

$$P(Y = 0 | X = x) = 1 - \eta(x) = \frac{1}{1 + \exp(x^t \beta)}$$

$$P(Y = y | X = x) = \sigma(z)^y (1 - \sigma(z))^{(1-y)}$$

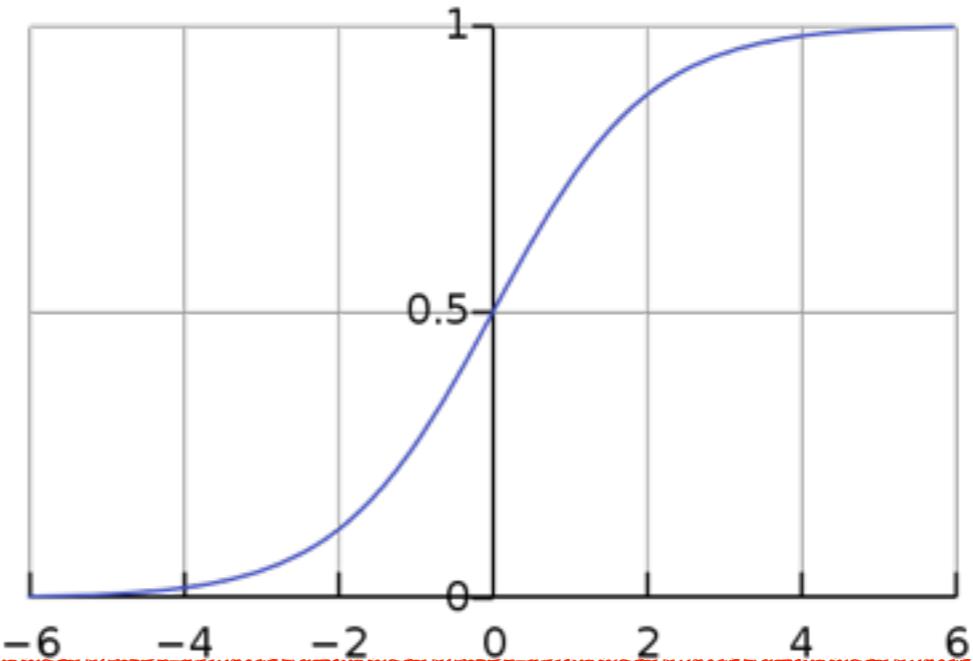
$$\sigma(z) = \frac{e^z}{1 + e^z}, \quad z = x^t \beta$$

$$\log P(Y = y | X = x) = y \cdot \log \sigma(z) + (1 - y) \cdot \log (1 - \sigma(z))$$

MLE for Binary Logistic Regression

$$\log \frac{\eta(x)}{1 - \eta(x)} = x^t \beta$$

Sigmoid function



$$P(Y = y | X = x) = \sigma(z)^y (1 - \sigma(z))^{(1-y)}$$

$$\sigma(z) = \frac{e^z}{1 + e^z}, \quad z = x^t \beta$$

$$\log P(Y = y | X = x) = y \cdot \log \sigma(z) + (1 - y) \cdot \log (1 - \sigma(z))$$

MLE for Binary Logistic Regression

$$P(Y = y|X = x) = \sigma(z)^y (1 - \sigma(z))^{(1-y)}$$

$$\sigma(z) = \frac{e^z}{1 + e^z}, \quad z = x^t \beta$$

$$\log P(Y = y|X = x) = y \cdot \log \sigma(z) + (1 - y) \cdot \log (1 - \sigma(z))$$

The Log-likelihood function is given by

$$\ell(\beta) = \sum_{i=1}^n y_i \cdot \log \sigma(z_i) + (1 - y_i) \cdot \log (1 - \sigma(z_i))$$

$$\frac{\partial \ell(\beta)}{\partial \beta} = ???$$

gradient (vector)

$$\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta} = ???$$

Hessian matrix

MLE for Binary Logistic Regression

$$P(Y = y|X = x) = \sigma(z)^y (1 - \sigma(z))^{(1-y)}$$

$$\sigma(z) = \frac{e^z}{1 + e^z}, \quad z = x^t \beta$$

$$\log P(Y = y|X = x) = y \cdot \log \sigma(z) + (1 - y) \cdot \log (1 - \sigma(z))$$

$$\begin{aligned}\frac{\partial}{\partial z} \sigma(z) &= \frac{\partial}{\partial z} \frac{1}{1 + e^{-z}} \\&= \frac{e^{-z}}{(1 + e^{-z})^2} = \frac{e^{-z}}{1 + e^{-z}} \frac{1}{1 + e^{-z}} \\&= \frac{1}{1 + e^z} \frac{e^z}{1 + e^z} = (1 - \sigma(z))\sigma(z)\end{aligned}$$

MLE for Binary Logistic Regression

$$P(Y = y|X = x) = \sigma(z)^y (1 - \sigma(z))^{(1-y)}$$

$$\sigma(z) = \frac{e^z}{1 + e^z}, \quad z = x^t \beta$$

$$\log P(Y = y|X = x) = \underline{y \cdot \log \sigma(z)} + (1 - y) \cdot \log (1 - \sigma(z))$$

$$\frac{\partial}{\partial \beta} \log P(Y = y|X = x)$$

$$= \frac{y}{\sigma(z)} \frac{\partial \sigma(z)}{\partial z} \frac{\partial z}{\partial \beta} - \frac{1 - y}{1 - \sigma(z)} \frac{\partial \sigma(z)}{\partial z} \frac{\partial z}{\partial \beta}$$

$$= \frac{y}{\sigma(z)} \sigma(z)(1 - \sigma(z))x - \frac{1 - y}{1 - \sigma(z)} \sigma(z)(1 - \sigma(z))x$$

$$= (y - \sigma(z))x$$

MLE for Binary Logistic Regression

$$P(Y = y|X = x) = \sigma(z)^y (1 - \sigma(z))^{(1-y)}$$

$$\sigma(z) = \frac{e^z}{1 + e^z}, \quad z = x^t \beta$$

$$\log P(Y = y|X = x) = \underline{y \cdot \log \sigma(z)} + (1 - y) \cdot \log (1 - \sigma(z))$$

$$\begin{aligned} & \frac{\partial}{\partial \beta} \log P(Y = y|X = x) \\ &= \frac{y}{\sigma(z)} \frac{\partial \sigma(z)}{\partial z} \frac{\partial z}{\partial \beta} - \frac{1 - y}{1 - \sigma(z)} \frac{\partial \sigma(z)}{\partial z} \frac{\partial z}{\partial \beta} \\ &= \frac{y}{\sigma(z)} \cancel{\sigma(z)(1 - \sigma(z))} x - \frac{1 - y}{1 - \cancel{\sigma(z)}} \sigma(z) \cancel{(1 - \sigma(z))} x \\ &= (y - \sigma(z))x \end{aligned}$$

MLE for Binary Logistic Regression

$$P(Y = y|X = x) = \sigma(z)^y (1 - \sigma(z))^{(1-y)}$$

$$\sigma(z) = \frac{e^z}{1 + e^z}, \quad z = x^t \beta$$

$$\log P(Y = y|X = x) = \underline{y \cdot \log \sigma(z)} + (1 - y) \cdot \log (1 - \sigma(z))$$

$$\begin{aligned} & \frac{\partial}{\partial \beta} \log P(Y = y|X = x) \\ &= \frac{y}{\sigma(z)} \frac{\partial \sigma(z)}{\partial z} \frac{\partial z}{\partial \beta} - \frac{1 - y}{1 - \sigma(z)} \frac{\partial \sigma(z)}{\partial z} \frac{\partial z}{\partial \beta} \\ &= \frac{y}{\sigma(z)} \cancel{\sigma(z)(1 - \sigma(z))} x - \frac{1 - \cancel{y}}{1 - \cancel{\sigma(z)}} \sigma(z) \cancel{(1 - \sigma(z))} x \\ &= (y - \sigma(z))x \end{aligned}$$

MLE for Binary Logistic Regression

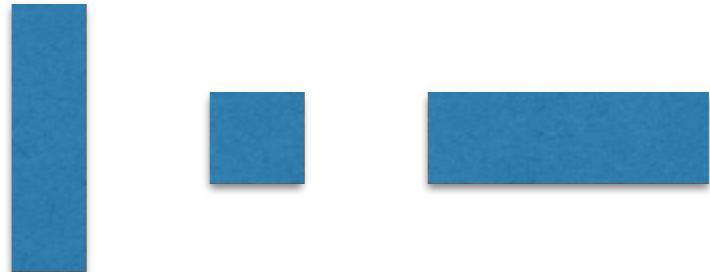
$$P(Y = y|X = x) = \sigma(z)^y (1 - \sigma(z))^{(1-y)}$$

$$\sigma(z) = \frac{e^z}{1 + e^z}, \quad z = x^t \beta$$

$$\log P(Y = y|X = x) = \underline{y \cdot \log \sigma(z)} + (1 - y) \cdot \log (1 - \sigma(z))$$

$$\boxed{\frac{\partial}{\partial \beta} \log P(Y = y|X = x) = (y - \sigma(z))x}$$

$$\frac{\partial^2}{\partial \beta \partial \beta} \log P(Y = y|X = x) = -x\sigma(z)(1 - \sigma(z))x^t$$

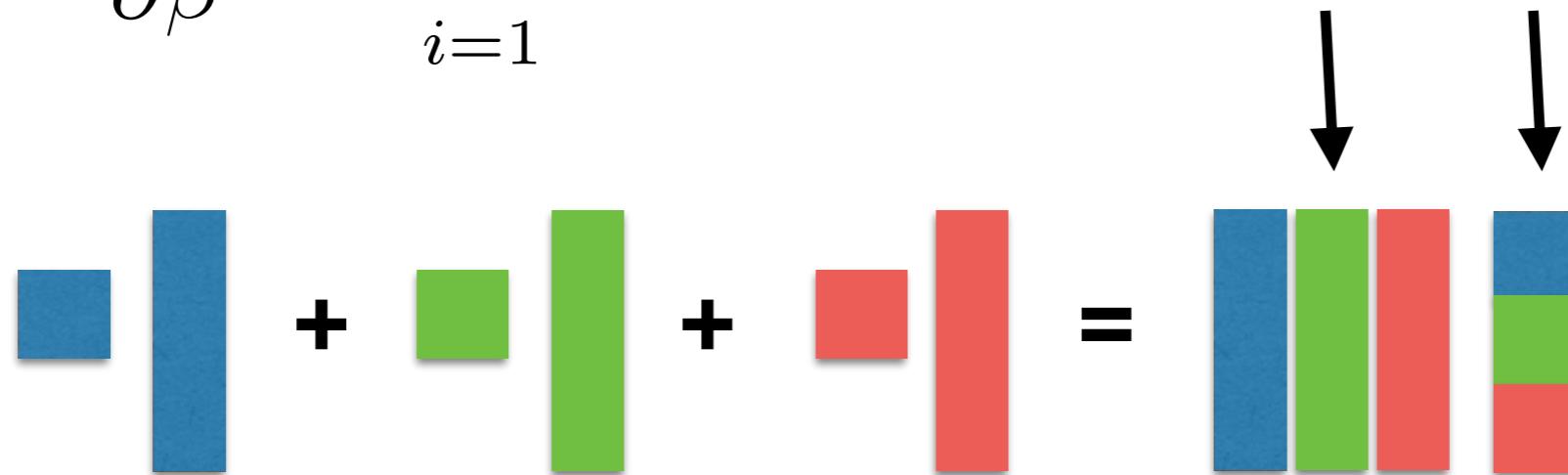


MLE for Binary Logistic Regression

The Log-likelihood function is given by

$$\ell(\beta) = \sum_{i=1}^n y_i \cdot \log \sigma(z_i) + (1 - y_i) \cdot \log (1 - \sigma(z_i))$$

$$\frac{\partial \ell(\beta)}{\partial \beta} = \sum_{i=1}^n (y_i - \sigma(z_i)) x_i = \mathbf{X}^t (\mathbf{y} - \mathbf{p})$$



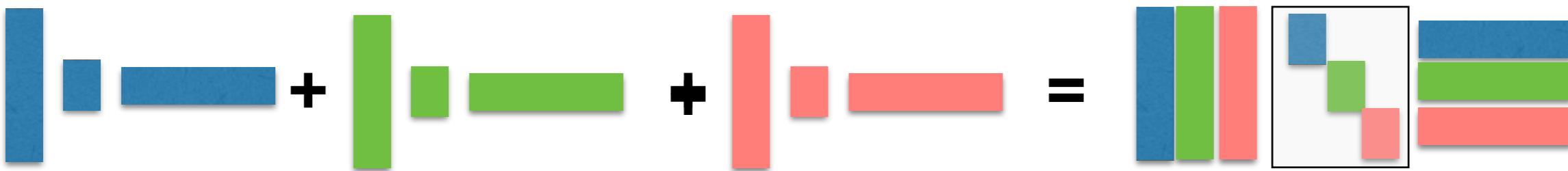
$$p_i = \sigma(z_i) = \frac{\exp(x_i^t \beta)}{1 + \exp(x_i^t \beta)} = \eta(x_i) = P(Y = 1 | X = x_i)$$

MLE for Binary Logistic Regression

The Log-likelihood function is given by

$$\ell(\beta) = \sum_{i=1}^n y_i \cdot \log \sigma(z_i) + (1 - y_i) \cdot \log (1 - \sigma(z_i))$$

$$\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta} = - \sum_{i=1}^n x_i \sigma(z_i)(1 - \sigma(z_i)) x_i^t = -\mathbf{X}^t \mathbf{W} \mathbf{X}$$



$$p_i = \sigma(z_i) = \frac{\exp(x_i^t \beta)}{1 + \exp(x_i^t \beta)} = \eta(x_i) = P(Y = 1 | X = x_i)$$

MLE for Binary Logistic Regression

The Log-likelihood function is given by

$$\ell(\beta) = \sum_{i=1}^n y_i \cdot \log \sigma(z_i) + (1 - y_i) \cdot \log (1 - \sigma(z_i))$$

$$\frac{\partial \ell(\beta)}{\partial \beta} = \mathbf{X}^t (\mathbf{y} - \mathbf{p})$$

gradient (vector)

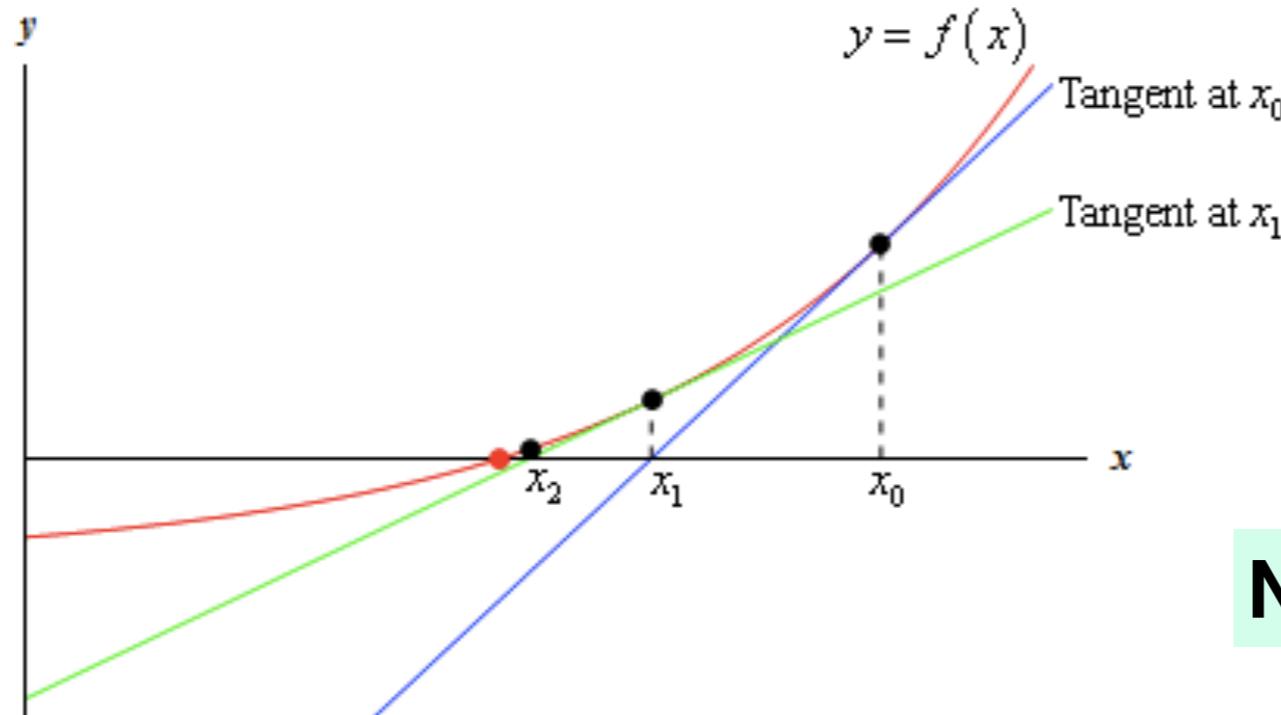
$$\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta} = -\mathbf{X}^t \mathbf{W} \mathbf{X}$$

Hessian matrix

Hessian is negative semi-definite, so log-likelihood function is **concave**, i.e., any local maximum is global maximum.

MLE is obtained via iterative **Newton-Raphson** to find the root of the derivative of log-likelihood function.

Newton-Raphson Method



New value

Old value

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) \implies x = x_0 - \frac{f(x_0)}{f'(x_0)}$$

$$\beta = \beta_0 + (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t (\mathbf{y} - \mathbf{p})$$

$$= (\mathbf{X}^t \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W} (\mathbf{X} \beta_0 + \mathbf{W}^{-1} (\mathbf{y} - \mathbf{p}))$$

Compute MLE for Logistic Regression

The MLE $\hat{\beta}$ can be obtained by the following [Reweighted LS Algorithm](#):

- Start with some initial values β^0
- Calculate the corresponding $p_i^0 = \eta^0(x_i)$ for $i = 1, \dots, n$; define

$$W = \text{diag}(p_i^0(1 - p_i^0))_{i=1}^n.$$

- Calculate

$$\mathbf{z} = X\beta^0 + W^{-1}(y - \mathbf{p}^0).$$

- Update $\beta^0 \leftarrow \beta^1$ with

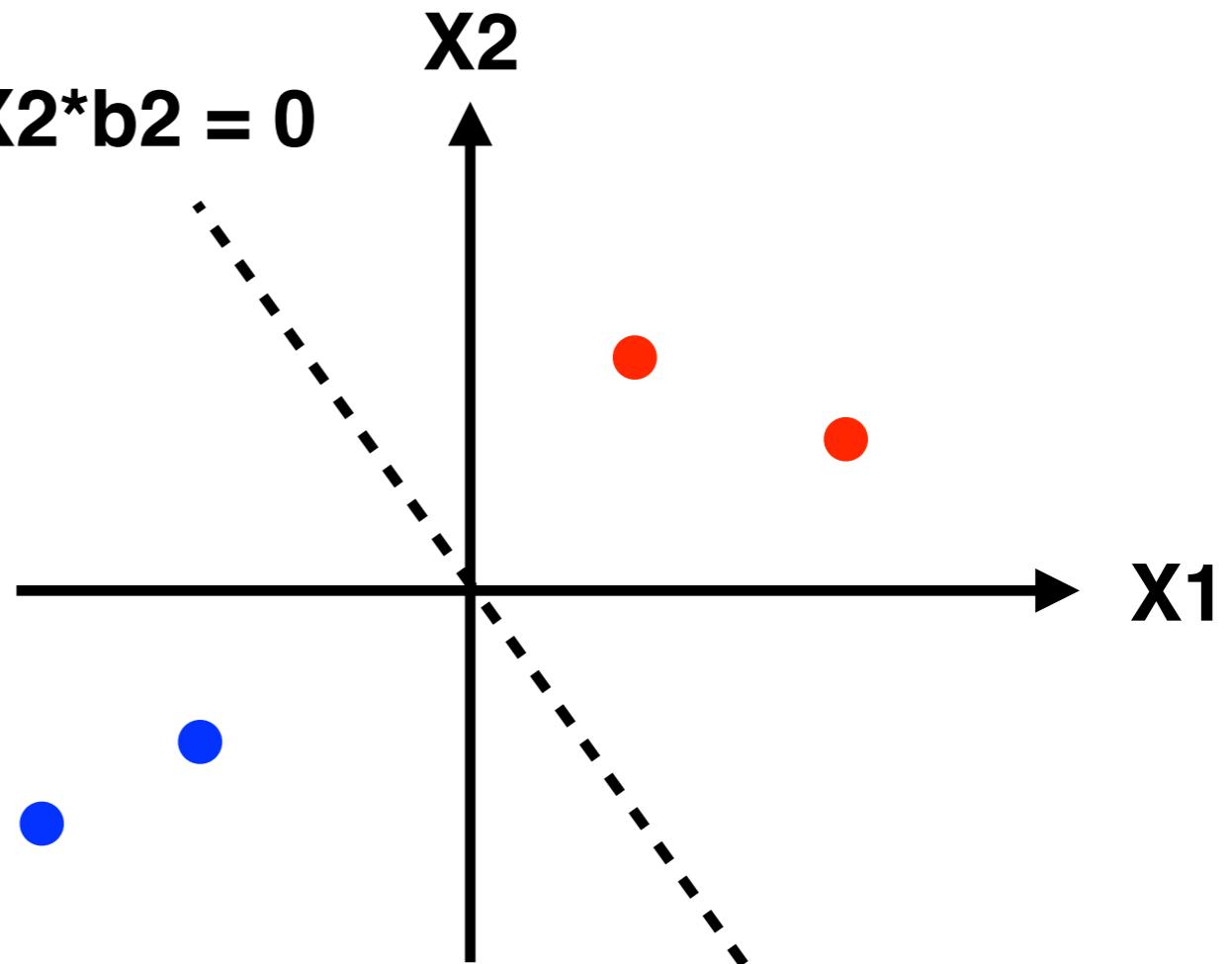
$$\beta^1 = (X^t W X)^{-1} X^t W \mathbf{z}.$$

and iterative the above steps until convergence.

Logistic Regression with Separable Data

$$X_1 \cdot b_1 + X_2 \cdot b_2 = 0$$

- $b_1=1, b_2=1$
- $b_1=10, b_2=10$
- $b_1=500, b_2=500$



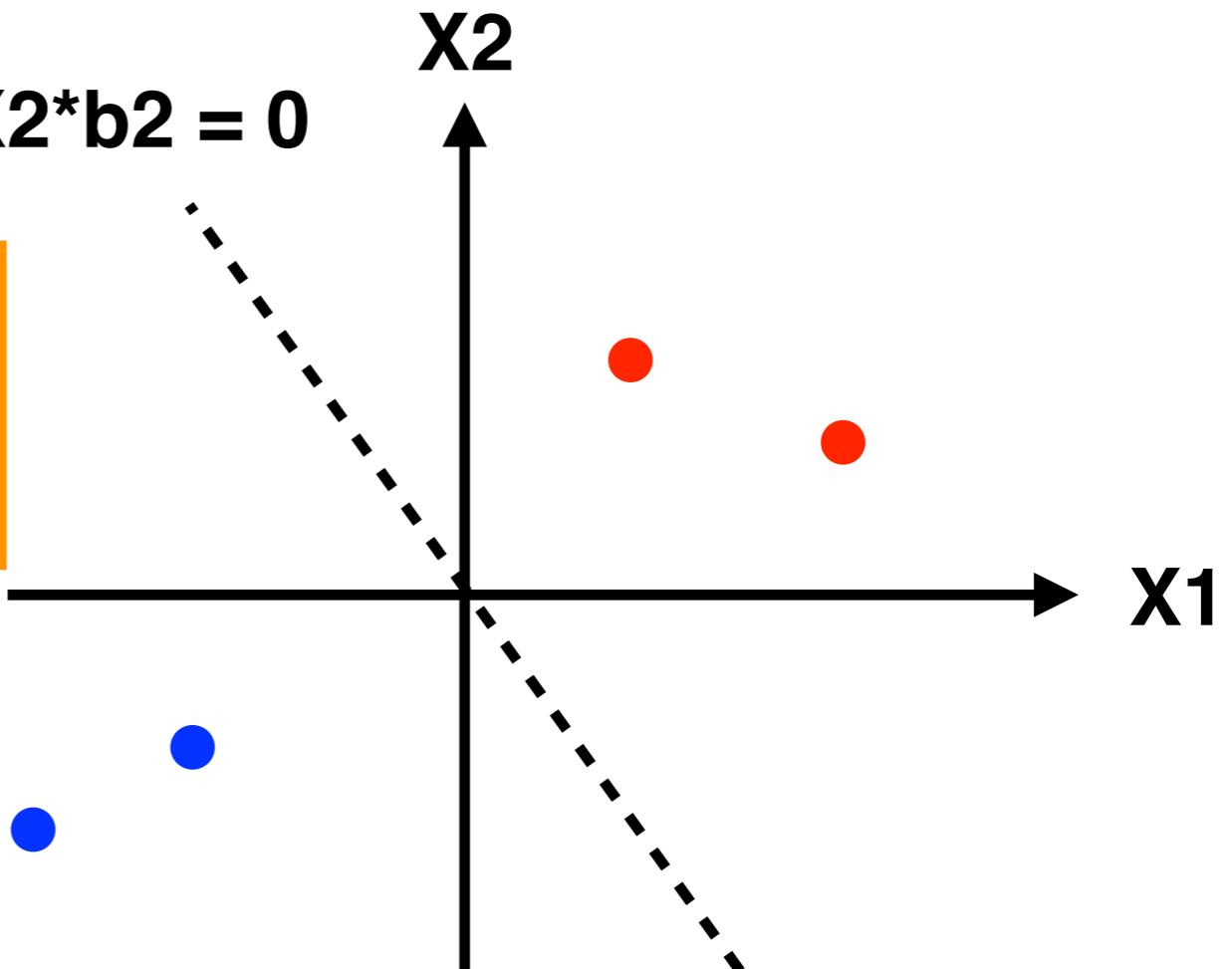
$$\frac{e^{(\beta_1 x_1 + \beta_2 x_2)}}{1 + e^{(\beta_1 x_1 + \beta_2 x_2)}} = \frac{1}{1 + e^{-(\beta_1 x_1 + \beta_2 x_2)}}$$

Logistic Regression with Separable Data

For separable data, the MLE doesn't converge. However, the separating line is well-defined.

- $b_1=1, b_2=1$
- $b_1=10, b_2=10$
- $b_1=500, b_2=500$

$$X_1 \cdot b_1 + X_2 \cdot b_2 = 0$$



$$\frac{e^{(\beta_1 x_1 + \beta_2 x_2)}}{1 + e^{(\beta_1 x_1 + \beta_2 x_2)}} = \frac{1}{1 + e^{-(\beta_1 x_1 + \beta_2 x_2)}}$$