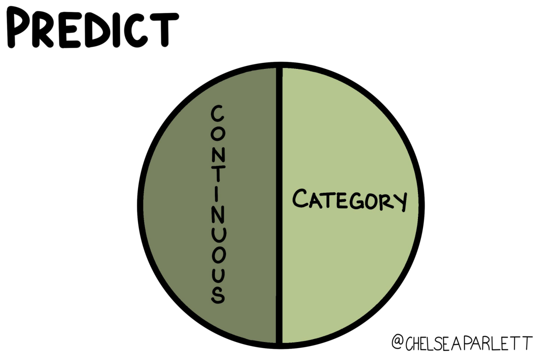


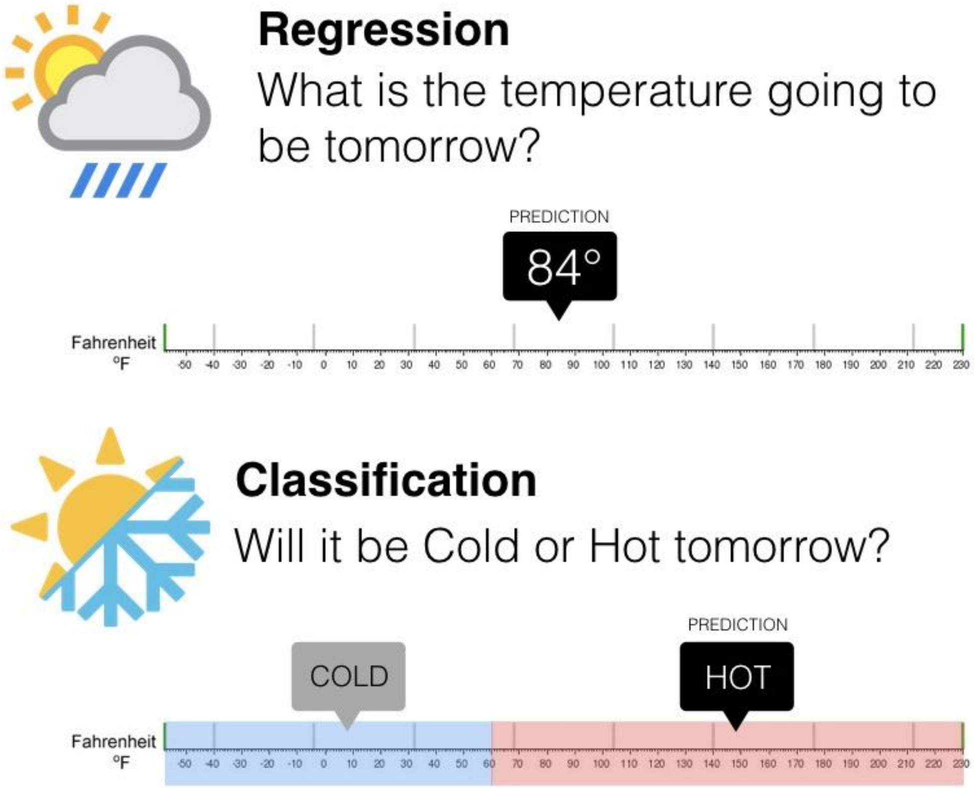
Logistic Regression I

Dr. Chelsea Parlett-Pelleriti

Outline

- Linear Regression in Disguise
- Linear Probability Models
- Link Functions
- Maximum Likelihood Estimation
- Loss Function

Regression vs. Classification



Linear Regression in Disguise

Linear Regression

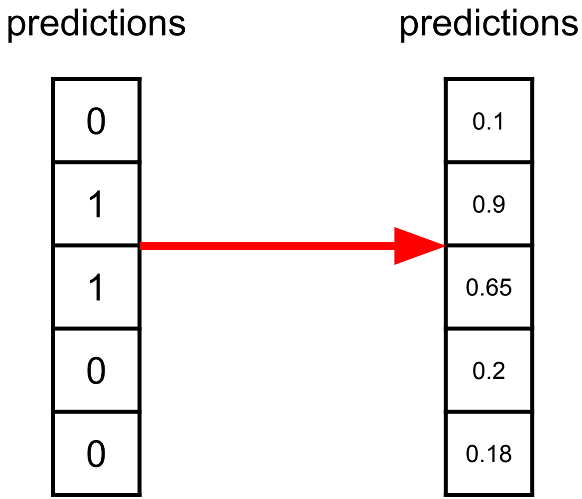
Continuous Variable (can be -∞ to ∞)

?

Logistic Regression

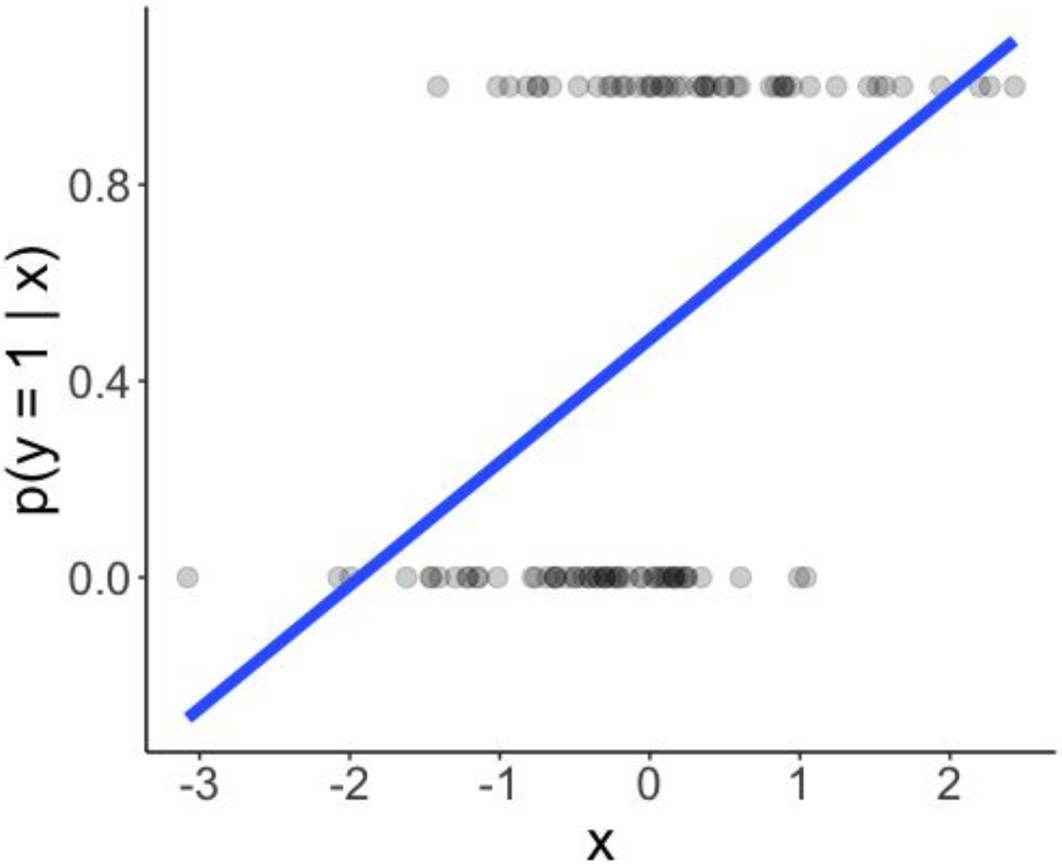
Binary Categorical Variable (can be 0 or 1)

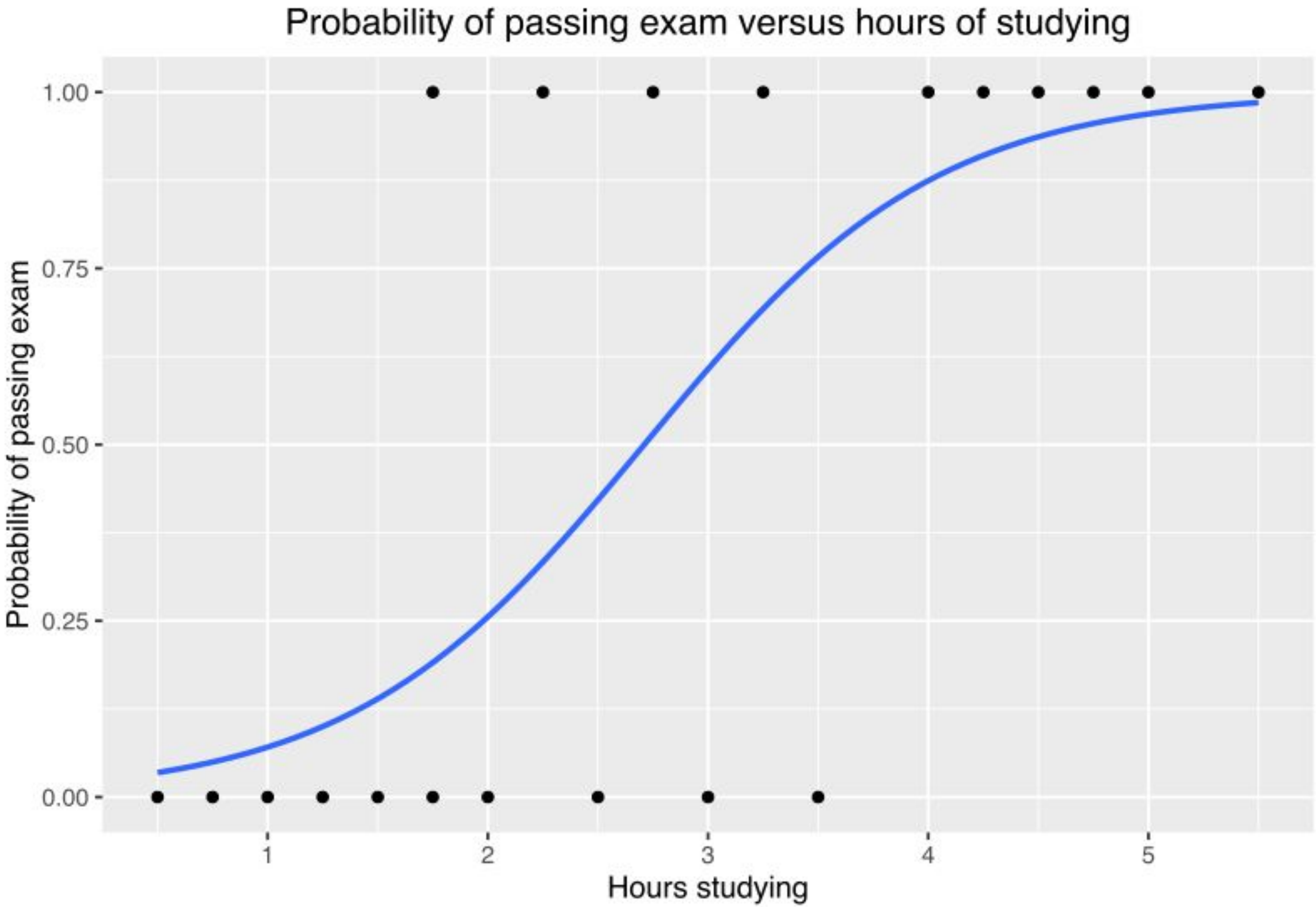
Probabilities



Linear Probability Model

$$p(y = 1 | x) = \beta_0 + \beta_1 * x$$





https://en.wikipedia.org/wiki/Logistic_regression

Link Functions

Link Functions

Two Ways to Use “Linear Regression” But Get Non-Linearity

1. Create New Features (Polynomial, GAMs…)
2. Link Functions ⭐

Link Functions

$$y = X\beta$$

$$y = g^{-1}(X\beta)$$

Link Functions

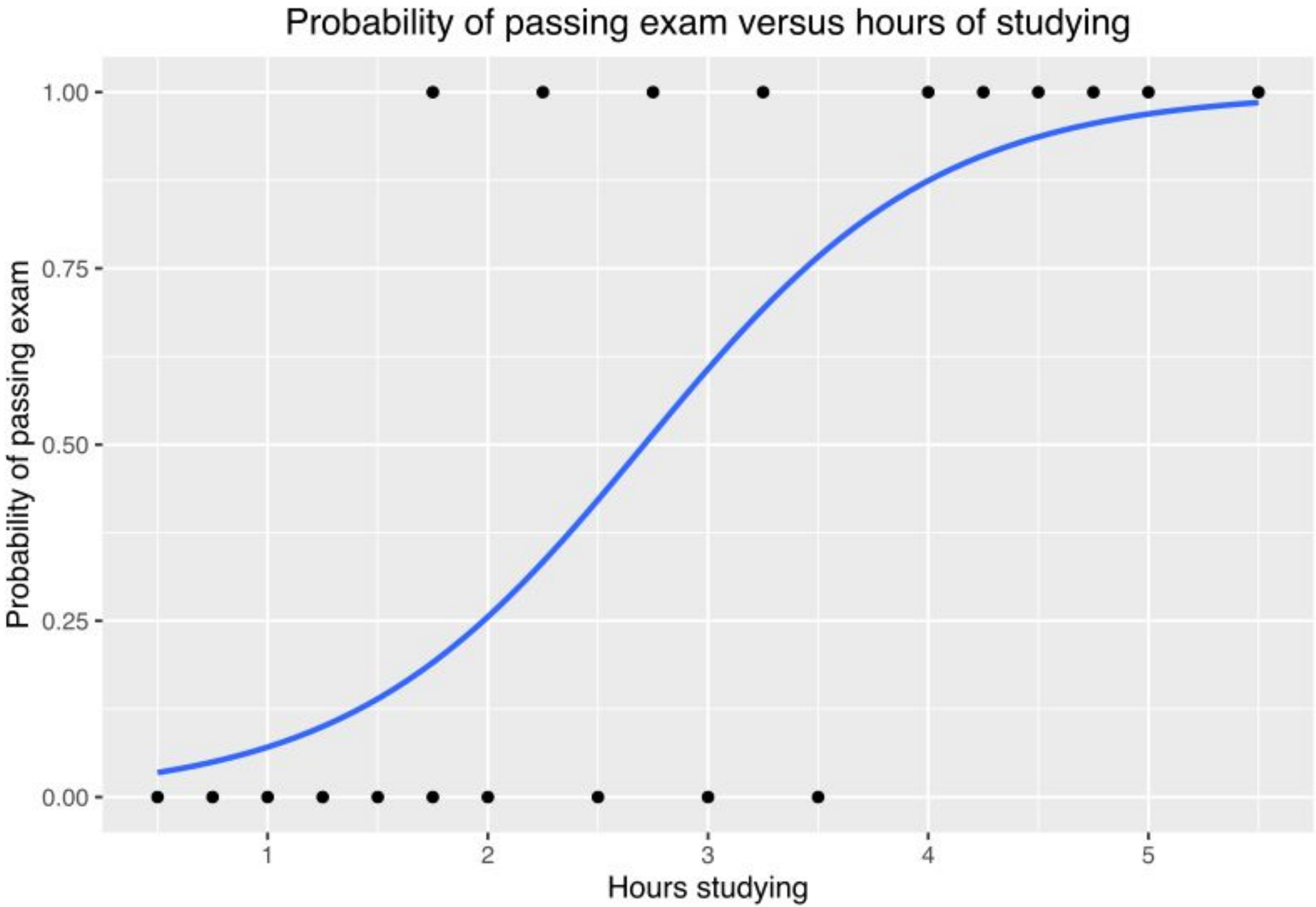
$$y = X\beta$$

$$y = g^{-1}(X\beta)$$

Link Functions

$$E(Y|X) = X\beta$$

$$E(Y|X) = g^{-1}(X\beta)$$



https://en.wikipedia.org/wiki/Logistic_regression

Probabilities, Odds, Log Odds

$$odds = \frac{p}{1-p}$$

Attempt 1: probabilities

Attempt 2: odds

$$\frac{p}{1-p} = \beta_0 + \beta_1 * x$$

Probabilities, Odds, Log Odds

$$odds = \frac{p}{1-p}$$

Attempt 1: probabilities

Attempt 2: odds

$$\frac{p}{1-p} = 0 + 1 * \text{dogs in stadium}$$

$$\text{log odds} = log(\frac{p}{1-p})$$

Probabilities, Odds, Log Odds

Attempt 1: probabilities

Attempt 2: odds

Attempt 3: log odds

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 * x$$

$$\text{log odds} = log(\frac{p}{1-p})$$

Probabilities, Odds, Log Odds

Attempt 1: probabilities

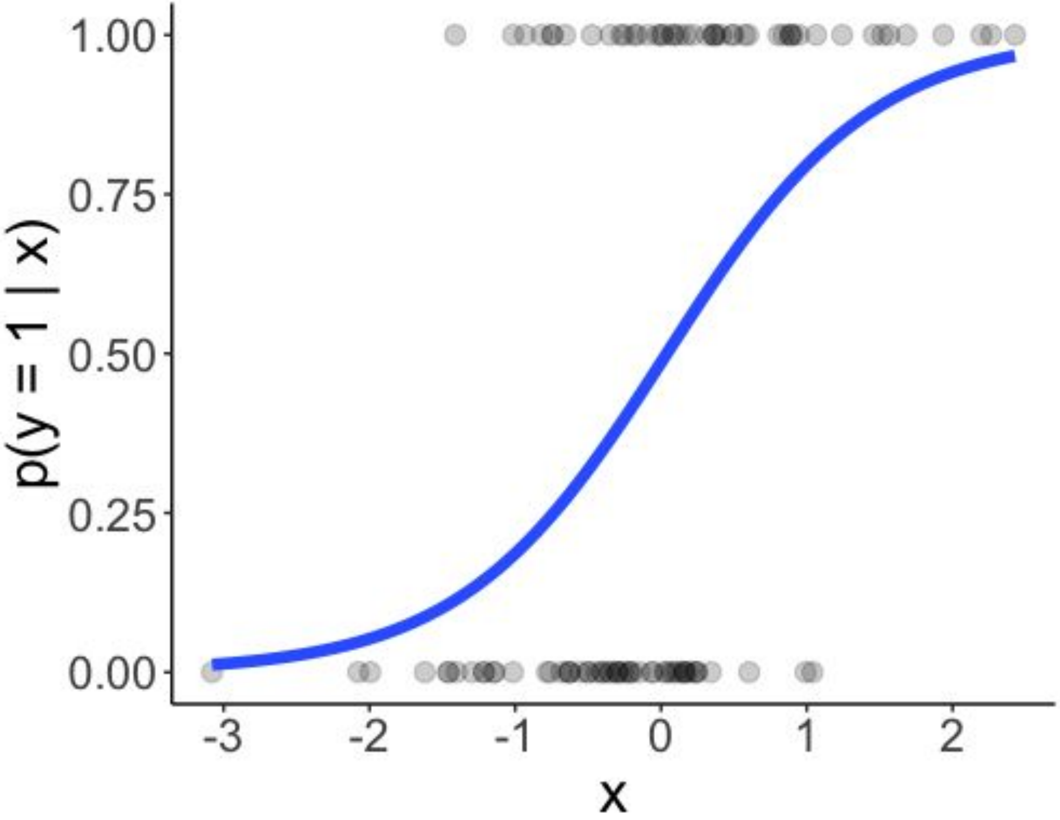
Attempt 2: odds

Attempt 3: log odds

$$log(\frac{p}{1-p}) = 0 + 1 * \text{dogs in stadium}$$

Logistic Regression

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * x$$



The Final Formula

$$log(\frac{p}{1-p}) = \beta_0 + \beta_1 * x_1$$

$$p = \frac{e^{\beta_0 + \beta_1 * x_1}}{1 + e^{\beta_0 + \beta_1 * x_1}}$$

The Final Formula

$$log \qquad * x_1$$

Do a little algebra to prove this to yourself!

$$p = \frac{e^{\beta_0 + \beta_1 * x_1}}{1 + e^{\beta_0 + \beta_1 * x_1}}$$

$$p = \frac{1}{1 + e^{-(\beta_0 + \beta_1 * x)}}$$

Link Functions (Again)

$$y = g^{-1}(X\beta)$$

General case

$$p = \frac{e^{X\beta}}{1 + e^{X\beta}}$$

Specific case

Our link function is

$$g(x) = log(\frac{x}{1-x})$$

which is the inverse of

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

Other Link Functions

Common distributions with typical uses and canonical link functions

Distribution	Support of distribution	Typical uses	Link name	Link function, $\mathbf{X}\beta = g(\mu)$	Mean function
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\beta = \mu$	$\mu = \mathbf{X}\beta$
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Negative inverse	$\mathbf{X}\beta = -\mu^{-1}$	$\mu = -(\mathbf{X}\beta)^{-1}$
Gamma					
Inverse Gaussian	real: $(0, +\infty)$		Inverse squared	$\mathbf{X}\beta = \mu^{-2}$	$\mu = (\mathbf{X}\beta)^{-1/2}$
Poisson	integer: $0, 1, 2, \ldots$	count of occurrences in fixed amount of time/space	Log	$\mathbf{X}\beta = \ln(\mu)$	$\mu = \exp(\mathbf{X}\beta)$
Bernoulli	integer: $\{0, 1\}$	outcome of single yes/no occurrence	Logit	$\mathbf{X}\beta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(\mathbf{X}\beta)}{1+\exp(\mathbf{X}\beta)} = \frac{1}{1+\exp(-\mathbf{X}\beta)}$
Binomial	integer: $0, 1, \ldots, N$	count of # of "yes" occurrences out of N yes/no occurrences		$\mathbf{X}\beta = \ln\left(\frac{\mu}{n-\mu}\right)$	
Categorical	integer: $[0, K)$	outcome of single K-way occurrence		$\mathbf{X}\beta = \ln\left(\frac{\mu}{1-\mu}\right)$	
	K-vector of integer: $[0, 1]$, where exactly one element in the vector has the value 1				
Multinomial	*K*-vector of integer: $[0, N]$	count of occurrences of different types (1 .. *K*) out of *N* total *K*-way occurrences			

From: https://en.wikipedia.org/wiki/Generalized_linear_model

Maximum Likelihood Estimation

Maximum Likelihood Estimation

$$\prod_{i;\ y_i=1} p(x_i) \qquad \prod_{i;\ y_i=0} 1 - p(x_i)$$

$$L(\beta_0, \beta_1) = \prod_{i;\ y_i=1} p(x_i) \ * \prod_{i;\ y_i=0} 1 - p(x_i)$$

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p(x_i)^{y_i} * (1 - p(x_i))^{1-y_i}$$

Maximum Likelihood Estimation

$$\prod_{i;\ y_i=0} 1 - p(x_i) \qquad \prod_{i;\ y_i=1} p(x_i)$$

$$L(\beta_0, \beta_1) = \prod_{i;\ y_i=1} p(x_i) \ * \prod_{i;\ y_i=0} 1 - p(x_i)$$

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p(x_i)^{y_i} * (1 - p(x_i))^{1-y_i}$$

Maximum Likelihood Estimation

$$\prod_{i;\ y_i=0} 1 - p(x_i) \qquad \prod_{i;\ y_i=1} p(x_i)$$

$$L(\beta_0, \beta_1) = \prod_{i;\ y_i=1} p(x_i) \ * \prod_{i;\ y_i=0} 1 - p(x_i)$$

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p(x_i)^{y_i} * (1 - p(x_i))^{1-y_i}$$

Maximum Likelihood Estimation

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p(x_i)^{y_i} * (1 - p(x_i))^{1-y_i}$$

$$l(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i * log(p(x_i)) + (1 - y_i) * log(1 - p(x_i))$$

Maximum Likelihood Estimation

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p(x_i)^{y_i} * (1 - p(x_i))^{1-y_i}$$

$$l(\beta_0, \beta_1) = \sum_{i=1}^{n} \boxed{y_i} * \boxed{log(p(x_i))} + \boxed{(1 - y_i)} * \boxed{log(1 - p(x_i))}$$

For full proof see: https://arunaddagatla.medium.com/maximum-likelihood-estimation-in-logistic-regression-f86ff1627b67

Maximum Likelihood Estimation

$$L(\beta_0, \beta_1) = \prod_{i=1}^{n} p(x_i)^{y_i} * (1 - p(x_i))^{1-y_i}$$

$$l(\beta_0, \beta_1) = \boxed{-\frac{1}{N}} \sum_{i=1}^{n} y_i * log(p(x_i)) + (1 - y_i) * log(1 - p(x_i))$$

Loss Function

that assesses performance, **smaller is better**

For full proof see: https://arunaddagatla.medium.com/maximum-likelihood-estimation-in-logistic-regression-f86ff1627b67

$$-\frac{1}{N}\sum_{i=1}^{n} y_i * log(p(x_i)) + (1-y_i) * log(1-p(x_i))$$

Loss Function