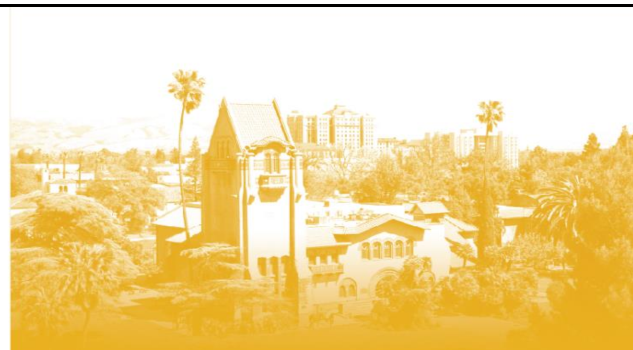
SJSU SAN JOSÉ STATE UNIVERSITY

# Vector Data – Classification
# Logistic Regression



1

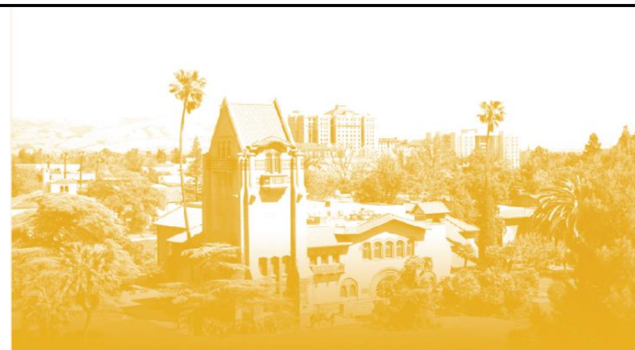
SJSU SAN JOSÉ STATE UNIVERSITY

## Agenda

- Introduction
- Mathematical Foundations
- Maximum Likelihood Estimation (MLE)
- Interpreting & Evaluating Model Coefficients

2

SJSU SAN JOSÉ STATE UNIVERSITY



**Logistic Regression**

3

SJSU SAN JOSÉ STATE UNIVERSITY

## Brief Recap of Logistic Regression Basics

- Logistic Regression is a statistical method for binary classification. Unlike linear regression, it predicts the probability of a binary outcome.

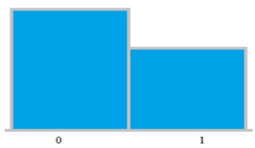
- Real World Examples:
  - disease presence
  - customer churn
  - loan defaults
  - spam detection

4

SJSU SAN JOSÉ STATE UNIVERSITY

## Discrete Distributions for Binary Cases

### Bernoulli Distribution

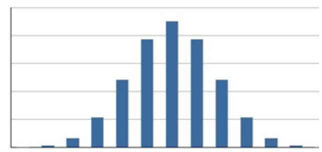



Y ~ Bern(p)

**Characteristics**
- Consists of a single trial
- 2 possible outcomes
- E(Y) = p
- Var(Y) = p × (1 – p)

**Uses**
- Guessing a single True/False question.

### Binomial Distribution



Y ~ B(n, p)

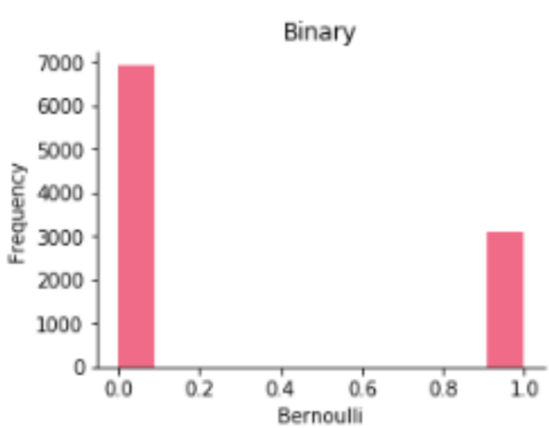
**Characteristics**
- Over the n trials, it measures the frequency of occurrence of one of the possible result.
- E(Y) = n × p
- P(Y = y) = C(y, n) × $p^y$ × $(1 – p)^{n-y}$
- Var(Y) = n × p × (1 – p)

**Uses**
- How many heads obtained if a coin is flipped a coin n times.
- Predict an event occur over a series of trials

5

SJSU SAN JOSÉ STATE UNIVERSITY

## Binary Data → Logistic Regression




**Data type:** binary
**Domain:** $0, 1$
**Examples:** True/False

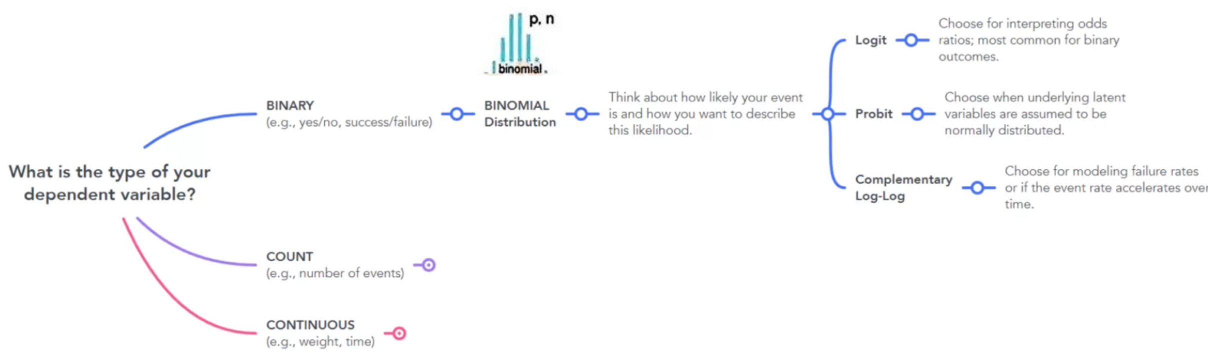
**Family:** `Binomial()`
**Link:** logit

**Model** = Logistic regression

6

SJSU SAN JOSÉ STATE UNIVERSITY

## GLM: Binary Variables, Binomial Distribution & Link Functions




https://statisticseasily.com/generalized-linear-model-distribution-and-link-function/

7

SJSU SAN JOSÉ STATE UNIVERSITY

## Summary Table of Different GLMs

| Response Variable Type | Suggested Distribution | Common Link Functions | Use Case |
|---|---|---|---|
| Binary Outcome (e.g., success/failure) | Binomial | Logit, Probit, Complementary Log-Log | Modeling probabilities of binary outcomes, such as presence/absence of a disease. |
| Count Data (e.g., number of events) | Poisson | Log, Identity, Square Root | Counting occurrences in fixed intervals, such as the number of calls received by a call center per hour. |
| Count Data with Overdispersion | Negative Binomial | Log, Identity | Count data that exhibit variability exceeding Poisson assumptions, such as the number of insurance claims per client. |
| Continuous Proportions | Beta | Logit, Probit | Proportions that vary between 0 and 1, such as the fraction of an area affected by a certain condition. |
| Positive Continuous Data | Gamma | Inverse, Log, Identity | Modeling waiting times or service times, where the response variable is always positive. |
| Normally Distributed Data | Normal (Gaussian) | Identity | Continuous outcomes that are symmetrically distributed, such as test scores or heights. |

8

## Logistic Regression

- Objective
  - Making a predictive model for classification
- Extension of a Linear Regression (think GLM)
  - When the output $Y$ is categorical.
- Classification
  - Classifying a new record, where its class is unknown, into one of the two classes, based on the values of its predictor variables $\boldsymbol{X}$.
- Feature Selection
  - Finding factors distinguishing between records in different classes in terms of their predictor variables $\boldsymbol{X}$, or "predictor profile" (odds).

9

## Linear Regression vs Logistic Regression

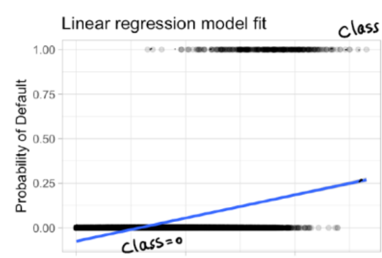
- Linear Regression (Prediction)
  - $Y$ : continuous value $(-\infty, +\infty)$

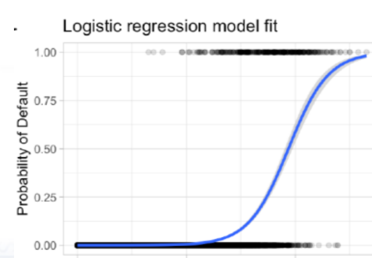
$$Y = \boldsymbol{X}^T\boldsymbol{\beta} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

$$Y|\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{X}^T\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$$

- Logistic Regression (Classification)
  - $Y$ : discrete value from M classes

$$\mathrm{P}(Y{=}C_j \,|\, \boldsymbol{x}; \boldsymbol{\beta}) \in [0,1] \text{ and } \Sigma_j \, \mathrm{P}(Y{=}C_j \,|\, \boldsymbol{x}; \boldsymbol{\beta}) = 1$$



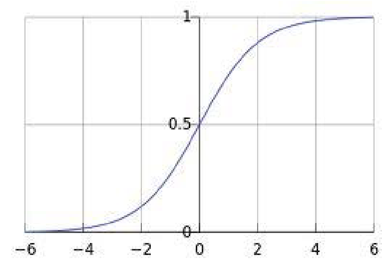




10

## Logistic Function

- Logistic Function / Sigmoid Function:
  - map any real-valued number $R$ into [0, 1]

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




- Note that the 1st derivative is simply:

$$\sigma'(x) = \left(\frac{1}{1+e^{-x}}\right)' = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}} = \sigma(x)\left(1-\sigma(x)\right)$$

P(1-P)

11

## Modeling Probabilities of Two Classes

- The probabilities of the 2 classes (0 and 1) are based on the logistic function $\sigma(x)$:

$$P(Y=1|\boldsymbol{X};\boldsymbol{\beta}) = \sigma(\boldsymbol{X}^T\boldsymbol{\beta}) = \frac{1}{1+e^{-\boldsymbol{X}^T\boldsymbol{\beta}}} = \frac{e^{\boldsymbol{X}^T\boldsymbol{\beta}}}{1+e^{\boldsymbol{X}^T\boldsymbol{\beta}}}$$

$$P(Y=0|\boldsymbol{X};\boldsymbol{\beta}) = 1 - P(Y=1|\boldsymbol{X};\boldsymbol{\beta}) = 1-\sigma(\boldsymbol{X}^T\boldsymbol{\beta}) = \frac{e^{-\boldsymbol{X}^T\boldsymbol{\beta}}}{1+e^{-\boldsymbol{X}^T\boldsymbol{\beta}}} = \frac{1}{1+e^{\boldsymbol{X}^T\boldsymbol{\beta}}}$$

- So, it's just the Bernoulli distribution:

$$Y|\boldsymbol{X} \sim \boldsymbol{Bern}(\sigma(\boldsymbol{X}^T\boldsymbol{\beta}))$$

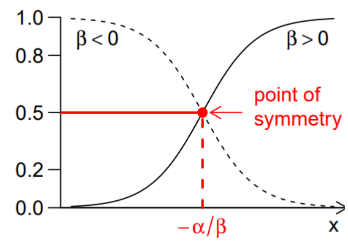
12

SJSU SAN JOSÉ STATE UNIVERSITY

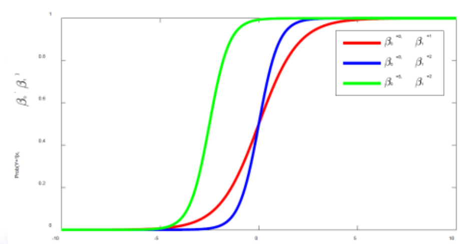
## 1D (One Variable) Example

- Here's a simple logistic regression model for a single explanatory variable $X_1$:

$$P(Y{=}1 \mid X_1, \beta_0, \beta_1) = \sigma(\beta_0 + \beta_1 X_1) = \frac{1}{1 + e^{-\beta_0 - \beta_1 X_1}}$$



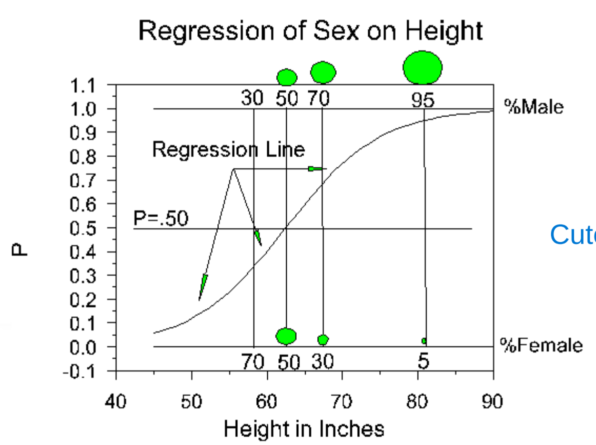

- What happens when $\beta_1$ increases?
- What does the coefficient $\beta_0$ represent?



13

SJSU SAN JOSÉ STATE UNIVERSITY

## Example




Cutoff=0.5

What do we know about $\beta_0$? Positive or negative?

14

SJSU SAN JOSÉ STATE UNIVERSITY

## Probability and Odds

- Probability (of class 1):

$$P(Y=1) = \frac{P(Y=1)}{P(Y=0)+P(Y=1)}$$

where

$$P(Y=1|\boldsymbol{X};\boldsymbol{\beta}) = \sigma(\boldsymbol{X}^T\boldsymbol{\beta}) = \frac{1}{1+e^{-\boldsymbol{X}^T\boldsymbol{\beta}}}$$

- Odds - Ratio of $P(Y=1)$ to $P(Y=0)$:

$$\text{Odds}(Y=1) = \frac{P(Y=1)}{P(Y=0)} = \frac{P(Y=1)}{1-P(Y=1)} \qquad \text{Odds} = \frac{P}{1-P} \quad \Rightarrow \quad P = \frac{\text{Odds}}{1-\text{Odds}}$$

15

SJSU SAN JOSÉ STATE UNIVERSITY

## Probability and Odds

- Odds

$$P(Y=1|\boldsymbol{X};\boldsymbol{\beta}) = \sigma(\boldsymbol{X}^T\boldsymbol{\beta}) = \frac{1}{1+e^{-\boldsymbol{X}^T\boldsymbol{\beta}}}$$

$$\text{Odds} = \frac{P}{1-P} = \frac{\dfrac{1}{1+e^{-\boldsymbol{X}^T\boldsymbol{\beta}}}}{1-\dfrac{1}{1+e^{-\boldsymbol{X}^T\boldsymbol{\beta}}}} = \frac{1}{e^{-\boldsymbol{X}^T\boldsymbol{\beta}}} = e^{\boldsymbol{X}^T\boldsymbol{\beta}}$$

$$\boxed{\text{logit}(P) = \log\left(\frac{P}{1-P}\right)}$$

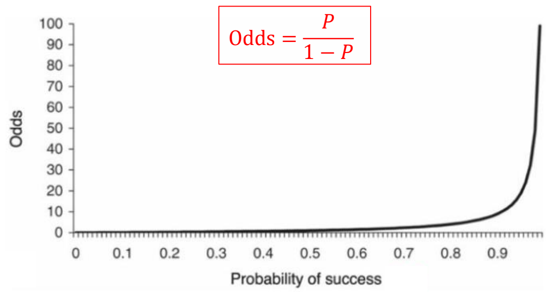
- Log Transformation of Odds $\log(\text{Odds}(Y=1))$ aka "Logit (Transformation)":

$$\underbrace{\log(\text{Odds}(Y=1))}_{\substack{\text{log odds} \\ \text{or} \\ \text{logit}(P)}} = \log\left(e^{\boldsymbol{X}^T\boldsymbol{\beta}}\right) = \boldsymbol{X}^T\boldsymbol{\beta} = \underbrace{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p}_{\text{linear combination of } \boldsymbol{X}}$$

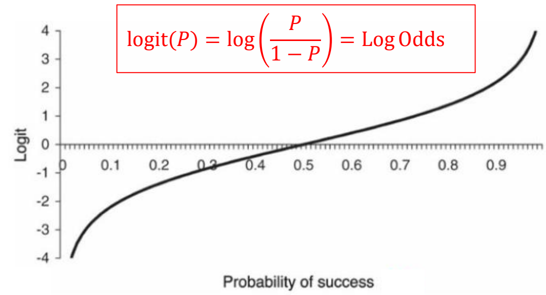
16

SJSU SAN JOSÉ STATE UNIVERSITY

## Relationship between Odds and Logit




$$\text{Odds} = \frac{P}{1-P}$$




$$\text{logit}(P) = \log\left(\frac{P}{1-P}\right) = \text{Log Odds}$$

logit function: $\mathrm{P}(\mathrm{Y}=1|\boldsymbol{X};\boldsymbol{\beta}) \Rightarrow \text{Odds}$

logistic function: $\sigma(\boldsymbol{X}^T\boldsymbol{\beta}) \in [0,1] \Rightarrow P(Y=1)$ using cutoff

17

SJSU SAN JOSÉ STATE UNIVERSITY

## Parameter Estimation with MLE

Maximum Likelihood Estimation (MLE)

- Given a dataset with $N$ data points
- For a single data object with predictors $\boldsymbol{X}_i$, and binary outcome $Y_i$
  - Let $p_i = P(Y_i = 1 \,|\boldsymbol{X}_i;\boldsymbol{\beta})$: the probability of $i$ in class 1
  - The probability of observing $Y_i$ would be:

$$p_i^{Y_i}(1-p_i)^{1-Y_i} \begin{cases} if\ Y_i = 1, then\ P_i \\ if\ Y_i = 0, then\ 1 - P_i \end{cases}$$

- Combining the two cases and include all datapoints ➔ Likelihood $L(\boldsymbol{\beta})$:

$$L(\boldsymbol{\beta}) = \prod_i p_i^{Y_i}(1-p_i)^{1-Y_i} = \prod_i \left(\frac{e^{\boldsymbol{X}_i{}^T\boldsymbol{\beta}}}{1+e^{\boldsymbol{X}_i{}^T\boldsymbol{\beta}}}\right)^{Y_i}\left(\frac{1}{1+e^{\boldsymbol{X}_i{}^T\boldsymbol{\beta}}}\right)^{1-Y_i}$$

18

SJSU SAN JOSÉ STATE UNIVERSITY

## Parameter Estimation with MLE

- To find the coefficients $\beta_i$, it's more straightforward to maximize the log likelihood:

$$\log L(\boldsymbol{\beta}) = \sum_i Y_i \log p_i + (1 - Y_i)\log(1 - p_i)$$

$$= \sum_i Y_i \log\left(\frac{e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}}{1 + e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}}\right) + (1 - Y_i)\log\left(\frac{1}{1 + e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}}\right)$$

$$= \sum_i Y_i \boldsymbol{X}_i^T\boldsymbol{\beta} - \log\left(1 + e^{\boldsymbol{X}_i^T\boldsymbol{\beta}}\right)$$

argmax of above

$$\frac{\partial \log L}{\partial \beta_j} = \sum_i (Y_j - p_j)X_{ij}$$

- Use gradient ascend update to compute $\beta_i$ numerically, :

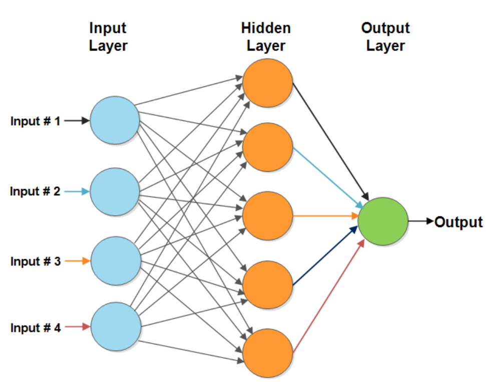
$$\beta_i^{new} = \beta_i^{\overset{old}{\cancel{new}}} + \eta \frac{\partial \log L}{\partial \beta_j}$$

19

SJSU SAN JOSÉ STATE UNIVERSITY

## Logistic Regression and Neural Network

- Basic Neural Network




| Activation function | Equation | Example | 1D Graph |
|---|---|---|---|
| Unit step (Heaviside) | $\phi(z) = \begin{cases} 0, & z < 0, \\ 0.5, & z = 0, \\ 1, & z > 0, \end{cases}$ | Perceptron variant | |
| Sign (Signum) | $\phi(z) = \begin{cases} -1, & z < 0, \\ 0, & z = 0, \\ 1, & z > 0, \end{cases}$ | Perceptron variant | |
| Linear | $\phi(z) = z$ | Adaline, linear regression | |
| Piece-wise linear | $\phi(z) = \begin{cases} 1, & z \geq \frac{1}{2}, \\ z + \frac{1}{2}, & -\frac{1}{2} < z < \frac{1}{2}, \\ 0, & z \leq -\frac{1}{2}, \end{cases}$ | Support vector machine | |
| Logistic (sigmoid) | $\phi(z) = \frac{1}{1 + e^{-z}}$ | Logistic regression, Multi-layer NN | |
| Hyperbolic tangent | $\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ | Multi-layer Neural Networks | |
| Rectifier, ReLU (Rectified Linear Unit) | $\phi(z) = max(0, z)$ | Multi-layer Neural Networks | |
| Rectifier, softplus | $\phi(z) = \ln(1 + e^z)$ | Multi-layer Neural Networks | |

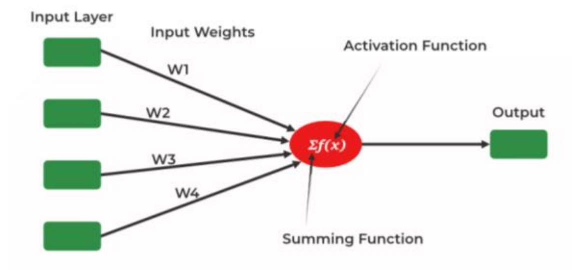


20

SJSU SAN JOSÉ STATE UNIVERSITY

## Logistic Regression and Neural Network

- Logistic Regression is similar to a single layer Neural Network:
  - Activation function ➔ Sigmoid function
  - Input nodes ➔ Predator variables $\boldsymbol{X}$
  - Weights ➔ Regression coefficients $\boldsymbol{\beta}$
- Different optimizers normally used:
  - MLE for Logistic Regression
  - Adam, BP, … for Neural Network
  - Linear model (LR) vs usually nonlinear (NN)




21

SJSU SAN JOSÉ STATE UNIVERSITY

**Interpreting Coefficients**

22

SJSU SAN JOSÉ STATE UNIVERSITY

## Model Coefficients

- The model coefficients $\widehat{\beta_i}$ are computed from the maximum likelihood.

```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  173
Model:                            GLM   Df Residuals:                      171
Model Family:                Binomial   Df Model:                            1
Link Function:                  logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -97.226
Date:                Thu, 26 Sep 2024   Deviance:                       194.45
Time:                        14:36:09   Pearson chi2:                     165.
No. Iterations:                     4   Pseudo R-squ. (CS):             0.1655
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -12.3508      2.629     -4.698      0.000     -17.503      -7.199
width          0.4972      0.102      4.887      0.000       0.298       0.697
==============================================================================
```

The intercept coefficient of −12.3508 denotes the baseline log odds exp(−12.3508) = 0.0004326 are the odds when *width* = 0

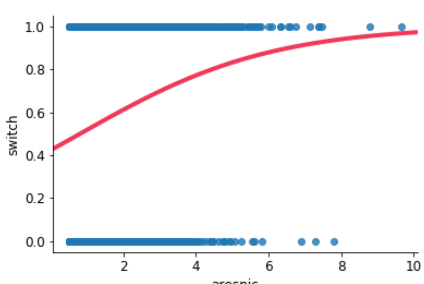
23

SJSU SAN JOSÉ STATE UNIVERSITY

## Model Coefficients

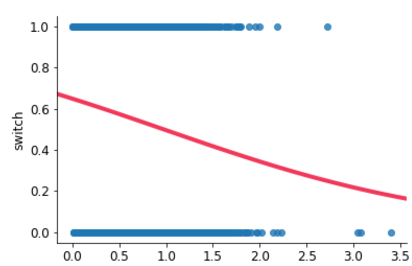
$\beta > 0$

- Ascending curve




$\beta < 0$

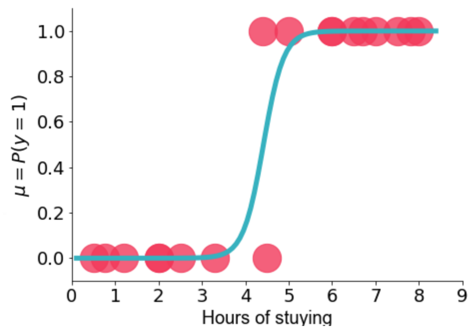
- Descending curve




24

SJSU SAN JOSÉ STATE UNIVERSITY

## Probability vs Logistic Fit




25

SJSU SAN JOSÉ STATE UNIVERSITY

## Log Odds Interpretation

- Logistic Model

$$\text{logit}(P) = \log\left(\frac{P}{1-P}\right) = \boldsymbol{X}^T\boldsymbol{\beta} = \beta_0 + \beta_1 X_1$$

- If $X_1$ is increased by one-unit

$$\text{logit}(P) = \log\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1(X_1 + 1)$$

$$\text{Odds} = \frac{P}{1-P} = e^{\beta_0+\beta_1(X_1+1)} = e^{\beta_0+\beta_1 X_1} e^{\beta_1}$$

odds are multiplied by $e^{\beta_1}$

26

## Profiling with Odds Ratio in Logistic Regression

Odds Ratio for a given predictor variable $X_i$ quantifies the change in odds of the outcome occurring for a one-unit increase in that predictor, holding all other variables constant.

- For $X_i$, the odds ratio (OR) is given as:

$$\mathrm{OR}(X_i) = \frac{\mathrm{odds}_{new}}{\mathrm{odds}_{orig}} = e^{\beta_i}$$

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -12.3508 | 2.629 | -4.698 | 0.000 | -17.503 | -7.199 |
| width | 0.4972 | 0.102 | 4.887 | 0.000 | 0.298 | 0.697 |

27

SJSU SAN JOSÉ STATE UNIVERSITY

## Standard Error (SE)

The standard error measures the variability or precision of an estimated coefficient. It helps in assessing the reliability of the coefficient estimates. It's given by:

$$\mathrm{SE}(\beta_i) = \sqrt{\mathrm{Var}(\widehat{\beta}_i)}$$

estimated variance of the coefficient

- It's used to construct confidence intervals and conduct hypothesis tests
- Smaller SE ➔ more precise estimate of $\beta_i$

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -12.3508 | 2.629 | -4.698 | 0.000 | -17.503 | -7.199 |
| width | 0.4972 | 0.102 | 4.887 | 0.000 | 0.298 | 0.697 |

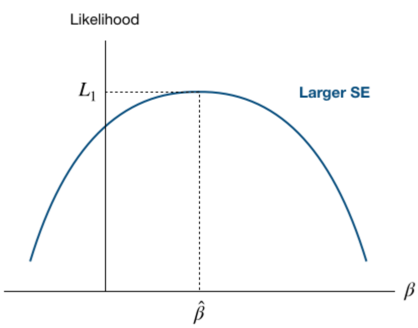
| | Intercept | width |
|---|---|---|
| Intercept | 6.910 | -0.267 |
| width | -0.267 | 0.010 |

covariance matrix

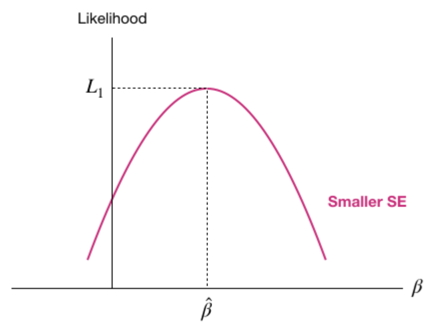
28

## Standard Error (SE)

- Flatter Peak
  - Maximum location harder to define
  - Larger SE




- Sharper Peak
  - Maximum location more clearly defined
  - Smaller SE




29

## Evaluation of Coefficients (z values)

The Z-statistic is used to test the significance of individual coefficients in a model:

$$z = \frac{\widehat{\beta}_i}{\mathrm{SE}(\widehat{\beta}_i)}$$

- Larger z ➔ $z \neq 0$ ➔ $\beta_i$ significant
- Rule of Thumb: cut-off value ➔ ~2

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -12.3508 | 2.629 | -4.698 | 0.000 | -17.503 | -7.199 |
| width | 0.4972 | 0.102 | 4.887 | 0.000 | 0.298 | 0.697 |

30

SJSU SAN JOSÉ STATE UNIVERSITY

## Evaluation of Coefficients (p values)

The p values can be used to determine the significance of individual coefficient $\beta_i$ :

Null Hypothesis $H_o$: $\beta_i = 0$ (no impact to outcome)

Alternate Hypothesis $H_1$: $\beta_i \neq 0$ (important)

- P-Value ≤ α ➔ reject the null hypothesis
  - This indicates that the predictor $X_i$ has a statistically significant relationship with outcome $Y$
- P-Value > α ➔ **cannot reject the null hypothesis**
  - This suggest insufficient evidence to conclude that the predictor variable $X_i$ is statistically associated with outcome $Y$

31

SJSU SAN JOSÉ STATE UNIVERSITY

## Example: Feature Selection using P Values

Feature Selection using the p values

Null Hypothesis $H_o$: $\beta_i = 0$ (no impact to outcome)

Alternate Hypothesis $H_1$: $\beta_i \neq 0$ (important)

- If 95% confidence level, $p < 0.05$
  ➔ **Reject $H_o$**
- If 99% confidence level, $p < 0.01$
  ➔ **Reject $H_o$**

| | Coef. | Std.Err. | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Age | -0.0359 | 0.0673 | -0.5340 | 0.5934 | -0.1678 | 0.0959 |
| Experience | 0.0450 | 0.0668 | 0.6740 | 0.5003 | -0.0859 | 0.1760 |
| Income | 0.0602 | 0.0030 | 20.2888 | 0.0000 | 0.0544 | 0.0660 |
| Family | 0.6182 | 0.0770 | 8.0239 | 0.0000 | 0.4672 | 0.7692 |
| CCAvg | 0.1634 | 0.0441 | 3.7078 | 0.0002 | 0.0770 | 0.2497 |
| Mortgage | 0.0007 | 0.0006 | 1.1961 | 0.2316 | -0.0005 | 0.0019 |
| SecuritiesAccount | -0.8701 | 0.3007 | -2.8938 | 0.0038 | -1.4595 | -0.2808 |
| CDAccount | 3.8389 | 0.3416 | 11.2393 | 0.0000 | 3.1695 | 4.5084 |
| Online | -0.7605 | 0.1657 | -4.5886 | 0.0000 | -1.0854 | -0.4357 |
| CreditCard | -1.0382 | 0.2131 | -4.8720 | 0.0000 | -1.4559 | -0.6205 |
| Education_Prof | 0.0987 | 0.1888 | 0.5226 | 0.6012 | -0.2714 | 0.4687 |
| Education_Under | -3.9654 | 0.2696 | -14.7084 | 0.0000 | -4.4938 | -3.4370 |
| Intercept | -8.3452 | 1.7916 | -4.6579 | 0.0000 | -11.8567 | -4.8337 |

32

SJSU SAN JOSÉ STATE UNIVERSITY

## Evaluation of Coefficients (Confidence Intervals)

The confidence intervals can be used to determine the uncertainly of individual coefficients:

$$[\hat{\beta}_i - z_{\alpha/2} \times \mathrm{SE}(\hat{\beta}_i), \hat{\beta}_i + z_{\alpha/2} \times \mathrm{SE}(\hat{\beta}_i)]$$

| Confidence Interval | z |
|---|---|
| 80% | 1.282 |
| 85% | 1.440 |
| 90% | 1.645 |
| 95% | 1.960 |
| 99% | 2.576 |
| 99.5% | 2.807 |
| 99.9% | 3.291 |

- For 95% confidence intervals for $\beta_i$:

$$[\hat{\beta}_i - 1.96 \times \mathrm{SE}(\hat{\beta}_i), \hat{\beta}_i + 1.96 \times \mathrm{SE}(\hat{\beta}_i)]$$

| | coef | std err | z | P>\|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | -12.3508 | 2.629 | -4.698 | 0.000 | -17.503 | -7.199 |
| width | 0.4972 | 0.102 | 4.887 | 0.000 | 0.298 | 0.697 |

95% CI

33

SJSU SAN JOSÉ STATE UNIVERSITY

## Example: Recap all the quantities

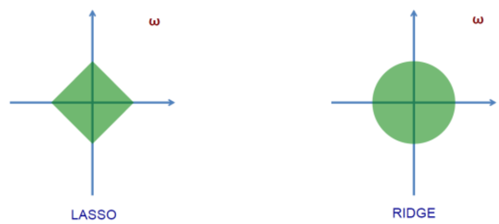
```
                 Generalized Linear Model Regression Results
==============================================================================
Dep. Variable:                      y   No. Observations:                  173
Model:                            GLM   Df Residuals:                      171
Model Family:                Binomial   Df Model:                            1
Link Function:                  logit   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -97.226
Date:                Thu, 26 Sep 2024   Deviance:                       194.45
Time:                        14:36:09   Pearson chi2:                     165.
No. Iterations:                     4   Pseudo R-squ. (CS):             0.1655
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -12.3508      2.629     -4.698      0.000     -17.503      -7.199
width          0.4972      0.102      4.887      0.000       0.298       0.697
==============================================================================
```

34

SJSU SAN JOSÉ STATE UNIVERSITY

## Regularization

- L1 (Lasso) Regularization
- L2 (Ridge) Regularization
- Elastic Net: Combining L1 and L2 regularization.






35

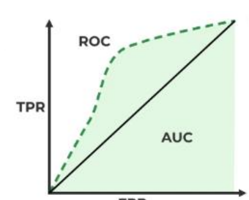
SJSU SAN JOSÉ STATE UNIVERSITY

## Model Evaluation

These are the standard model evaluation metrics for classification problems:

- Confusion Matrix
  - Summarizes the performance of the classification model.
- ROC Curve
  - Plots the true positive rate vs the false positive rate at various threshold/cutoff settings.
- AUC (Area Under the Curve)
  - Measures the overall performance of the model.

| | | Predicted 0 | Predicted 1 |
|---|---|---|---|
| Actual | 0 | TN | FP |
| | 1 | FN | TP |






36

SJSU SAN JOSÉ STATE UNIVERSITY

## Summary

- Logistic Regression is a powerful tool for binary classification problems, providing interpretable results through odds ratios and p-values.
- Formal mathematical formulation including using the MLE to estimate parameters.
- Logistic regression is similar to a single layer neural network with sigmoid function as the activation function.
- Hypothesis testing can be used to determine the significance of individual coefficients.
- $L_1$, $L_2$ and Elastic Net regularization can be used with logistic regression.
- Confusion matrix, ROC and AUC are popular model evaluation metrics for logistic regression.

37