Artificial Intelligence Club Week 7 Slides

November 12th, 2025

Supervised Learning Basics

Goal: learn a mapping from features (inputs) to labels (outputs) using labeled examples

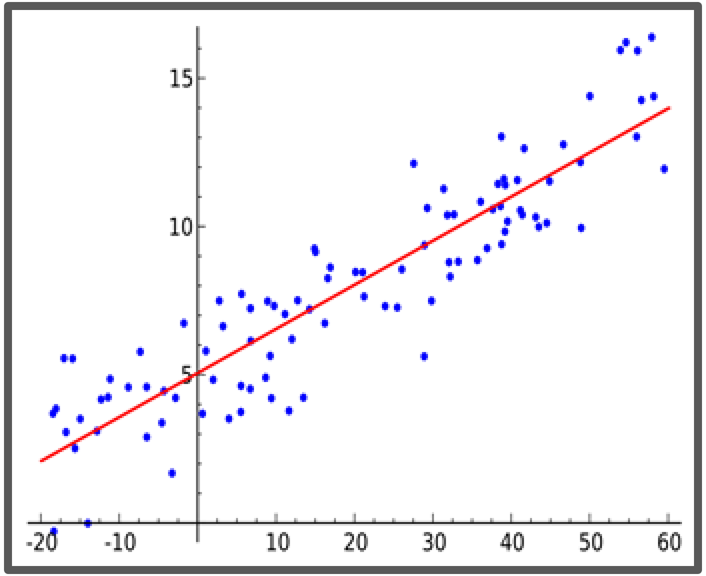
Training data: pairs of (features, target) used to fit a model

Choose a model and a loss function; minimize average loss on the training set

Tasks: regression (numbers) and classification (categories)

Metrics: Root mean squared error for regression or log loss for classification

Linear Regression



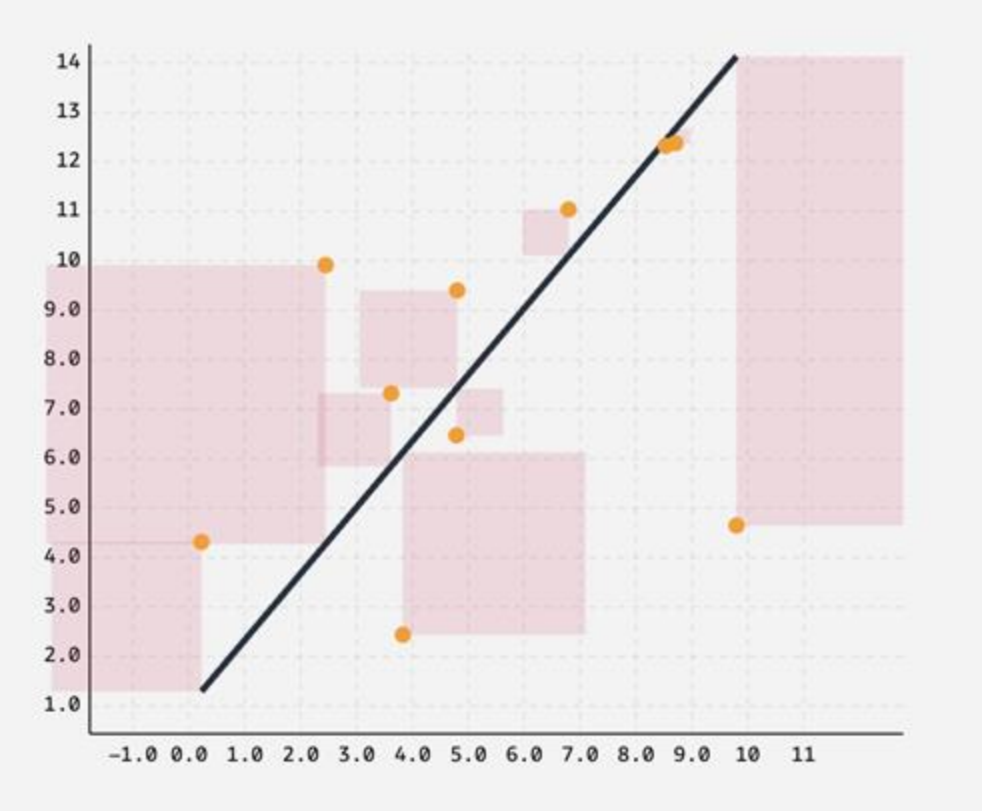
Model: prediction = weighted sum of features + bias

MLE/Ordinary Least Squares: choose coefficients that minimize the sum of squared errors

Assumptions: linear relationship and independent errors

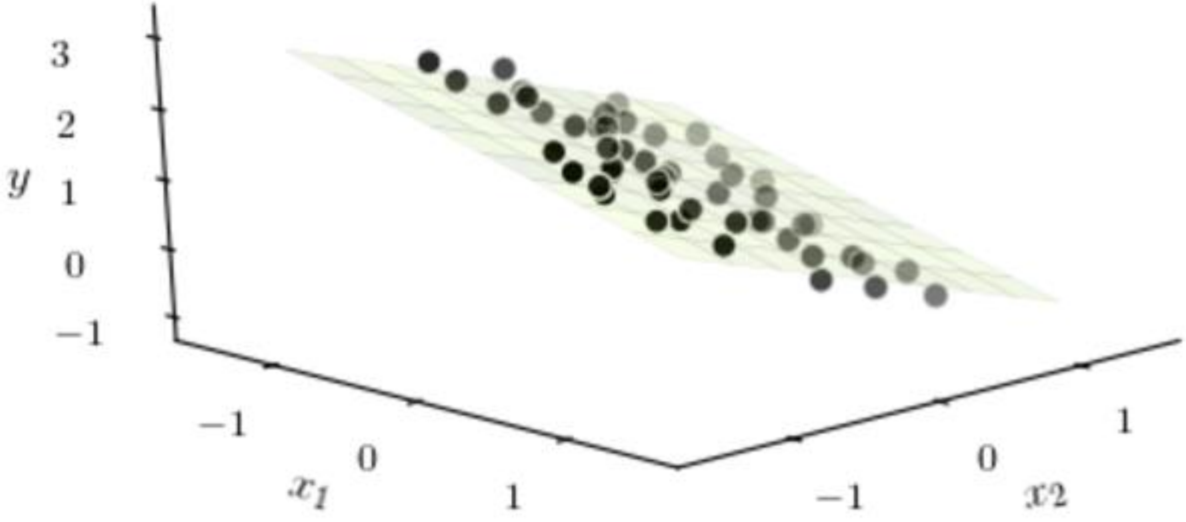
Evaluation: root mean squared error

- We minimize MSE (our loss function)
- We report RMSE (to put it in the same units as “y”)



Linear Regression

Linear Regression



Linear Regression Practice

https://www.mladdict.com/linear-regression-simulator

https://mlu-explain.github.io/linear-regression/

https://kenndanielso.github.io/mlrefined/blog_posts/8_Linear_regression/8_1_Least_squares_regression.html

Linear Regression

$$\mathbf{x}^{(1)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \ \mathbf{x}^{(2)} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \ \mathbf{x}^{(3)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}, \ y^{(1)} = 5, \ y^{(2)} = 6, \ y^{(3)} = 7$$

$$\hat{y}^{(i)} = \mathbf{w}^T \mathbf{x}^{(i)} + b, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \ b \in R$$

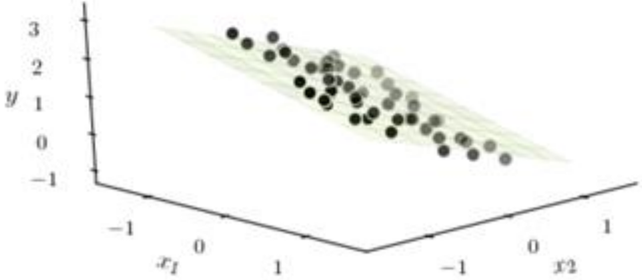
$$\mathbf{w} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad b = 0$$

$$\hat{y}^{(1)} = 0, \ \hat{y}^{(2)} = 0, \ \hat{y}^{(3)} = 0$$

$$\text{MSE} = \frac{1}{3} \sum_{i=1}^{3} (y^{(i)} - \hat{y}^{(i)})^2$$

$$\text{MSE} = \frac{1}{3} \left((5-0)^2 + (6-0)^2 + (7-0)^2 \right) = \frac{110}{3} \approx 36.67$$

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{110}{3}} \approx 6.06$$



Linear Regression

$$\frac{\partial \mathrm{MSE}}{\partial \mathbf{w}} = -\frac{2}{3}\sum_{i=1}^{3}(y^{(i)} - \hat{y}^{(i)})\mathbf{x}^{(i)}, \quad \frac{\partial \mathrm{MSE}}{\partial b} = -\frac{2}{3}\sum_{i=1}^{3}(y^{(i)} - \hat{y}^{(i)})$$

$$\mathbf{w}_{\text{new}} = \mathbf{w}_{\text{old}} - \eta\frac{\partial \mathrm{MSE}}{\partial \mathbf{w}}, \quad b_{\text{new}} = b_{\text{old}} - \eta\frac{\partial \mathrm{MSE}}{\partial b}$$

$$\frac{\partial \mathrm{MSE}}{\partial w_1} = -\frac{2}{3}\big((5-0)(1) + (6-0)(2) + (7-0)(3)\big) = -25.33$$

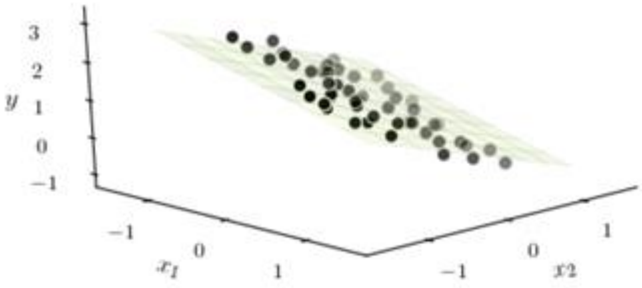
$$\frac{\partial \mathrm{MSE}}{\partial w_2} = -\frac{2}{3}\big((5-0)(2) + (6-0)(0) + (7-0)(1)\big) = -11.33$$

$$\frac{\partial \mathrm{MSE}}{\partial b} = -\frac{2}{3}\big((5-0) + (6-0) + (7-0)\big) = -12$$

$$w_1^{\text{new}} = w_1^{\text{old}} - \eta\frac{\partial \mathrm{MSE}}{\partial w_1} = 0 - 0.01 * (-25.33) = 0.2533$$

$$w_2^{\text{new}} = w_2^{\text{old}} - \eta\frac{\partial \mathrm{MSE}}{\partial w_2} = 0 - 0.01 * (-11.33) = 0.1133$$

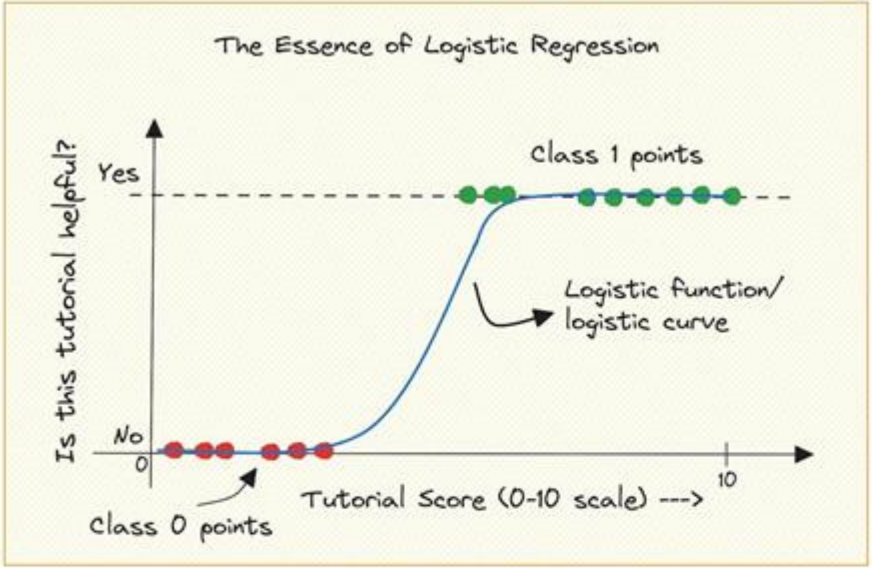
$$b^{\text{new}} = b^{\text{old}} - \eta\frac{\partial \mathrm{MSE}}{\partial b} = 0 - 0.01 * (-12) = 0.12$$



Logistic Regression

Model: probability of the positive class = sigmoid(weighted sum of features)

Fitting: maximize the log-likelihood (or minimize log loss) using gradient-based optimization



Outputs: 0-1 predicted probabilities; choose a threshold (often halfway) to produce class labels

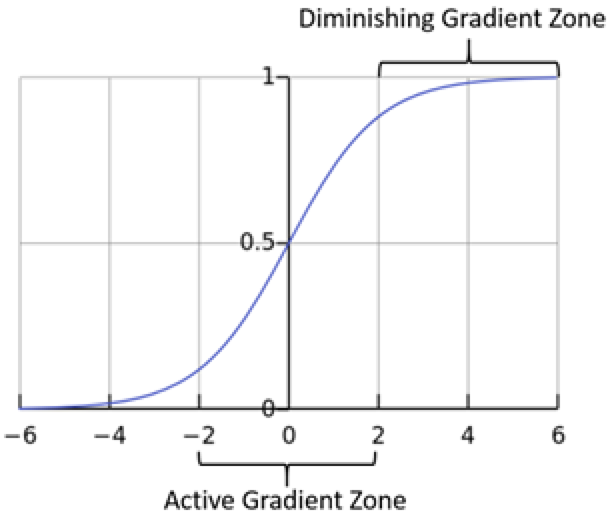
Evaluation: accuracy and log loss

Logistic Regression Practice

https://mlu-explain.github.io/logistic-regression/

https://kenndanielso.github.io/mlrefined/blog_posts/9_Linear_twoclass_classification/9_1_Logistic_regression.html

$$A = \frac{1}{1+e^{-x}}$$



Note:
After we are done (our loss looks low enough), we convert our percentage to **Class 0** or **Class 1**.

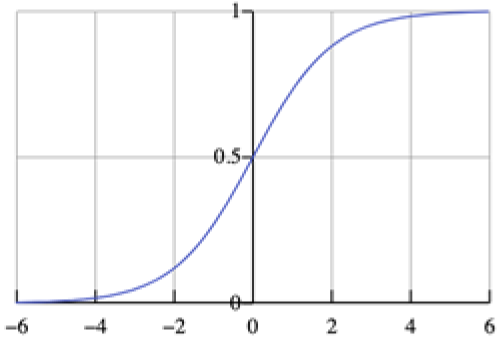
Logistic Regression

$$\mathbf{x}^{(1)} = \begin{bmatrix} 1 \\ 2 \end{bmatrix},\ \mathbf{x}^{(2)} = \begin{bmatrix} 2 \\ 0 \end{bmatrix},\ \mathbf{x}^{(3)} = \begin{bmatrix} 3 \\ 1 \end{bmatrix},\ y^{(1)} = 0,\ y^{(2)} = 1,\ y^{(3)} = 1$$

$$\hat{y}^{(i)} = \sigma(\mathbf{w}^T \mathbf{x}^{(i)} + b), \quad \sigma(z) = \frac{1}{1 + e^{-z}}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix},\ b \in R$$

$$\mathbf{w} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad b = 0$$

$$\hat{y}^{(1)} = \hat{y}^{(2)} = \hat{y}^{(3)} = \sigma(0) = 0.5$$



$$\text{Binary Cross-Entropy Loss: BCE} = -\frac{1}{3} \sum_{i=1}^{3} \left[y^{(i)} \log \hat{y}^{(i)} + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \right]$$

$$\text{BCE} = -\frac{1}{3} \left[((0) \log 0.5 + (1-0) \log 0.5) + ((1) \log 0.5 + (1-1) \log 0.5) + ((1) \log 0.5 + (1-1) \log 0.5) \right] = 0.693$$

Logistic Regression

$$\frac{\partial \mathrm{BCE}}{\partial \mathbf{w}} = \frac{1}{3}\sum_{i=1}^{3}(\hat{y}^{(i)} - y^{(i)})\mathbf{x}^{(i)}, \quad \frac{\partial \mathrm{BCE}}{\partial b} = \frac{1}{3}\sum_{i=1}^{3}(\hat{y}^{(i)} - y^{(i)})$$

$$\frac{\partial \mathrm{BCE}}{\partial w_1} = \frac{1}{3}\big((0.5-0)*1+(0.5-1)*2+(0.5-1)*3\big) = \frac{1}{3}(0.5-1-1.5) = -0.667$$

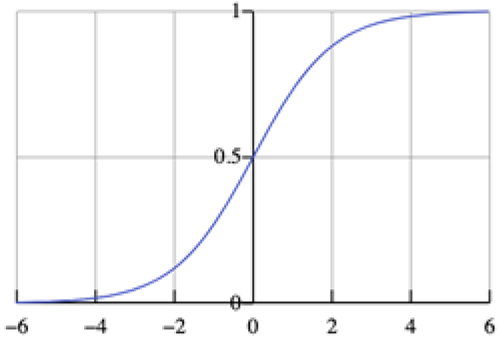
$$\frac{\partial \mathrm{BCE}}{\partial w_2} = \frac{1}{3}\big((0.5-0)*2+(0.5-1)*0+(0.5-1)*1\big) = \frac{1}{3}(1+0-0.5) = 0.167$$

$$\frac{\partial \mathrm{BCE}}{\partial b} = \frac{1}{3}\big((0.5-0)+(0.5-1)+(0.5-1)\big) = \frac{1}{3}(0.5-0.5-0.5) = -0.167$$

$$w_1^{\text{new}} = w_1^{\text{old}} - \eta\frac{\partial \mathrm{BCE}}{\partial w_1} = 0 - 0.01*(-0.667) = 0.00667$$

$$w_2^{\text{new}} = w_2^{\text{old}} - \eta\frac{\partial \mathrm{BCE}}{\partial w_2} = 0 - 0.01*(0.167) = -0.00167$$

$$b^{\text{new}} = b^{\text{old}} - \eta\frac{\partial \mathrm{BCE}}{\partial b} = 0 - 0.01*(-0.167) = 0.00167$$



Feedback

Hackathon
Interest Form!

Connections + Pizza + Talk to Officers