Model Evaluation

7c

CS 3244
Machine Learning

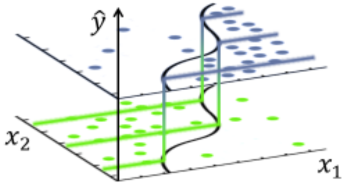
NUS National University of Singapore | Computing

Recap from Week 07b

Classification

$y \in \{0,1\}$ binary

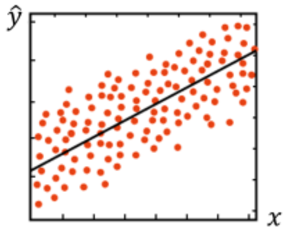
$y \in \{y_A, y_B, \ldots\}$ multi-class



$y = M(\boldsymbol{x}), \quad \boldsymbol{x} = \vec{x} = (x_1, x_2)^\top$

Regression

$y \in \mathbb{R}$ any real number



$y = M(x), \quad x = x_1$

Image credit:
https://www.javatpoint.com/regression-vs-classification-in-machine-learning

Student Learning Outcomes

What did we learn this week?

1. Recap: Classification vs. Regression
2. Classification Metrics
   1. Accuracy
   2. Confusion Matrix, TP, TN, FP, FN
   3. Precision, Recall, F_1
   4. ROC, AUC
   5. Micro- and Macro-Averaging
   6. PR-AUC (Average Precision)
3. Regression Metrics [W08a]

Performance Metrics

Week 07b: Lecture Outline

1. Recap: Supervised learning Classification vs. Regression
2. Classification Metrics
3. Regression Metrics
4. Unsupervised learning metrics [W11]

Regression Evaluation Metrics

Department of Computer Science
School of Computing

Classification

$\hat{y} \in \{0,1\}$ binary

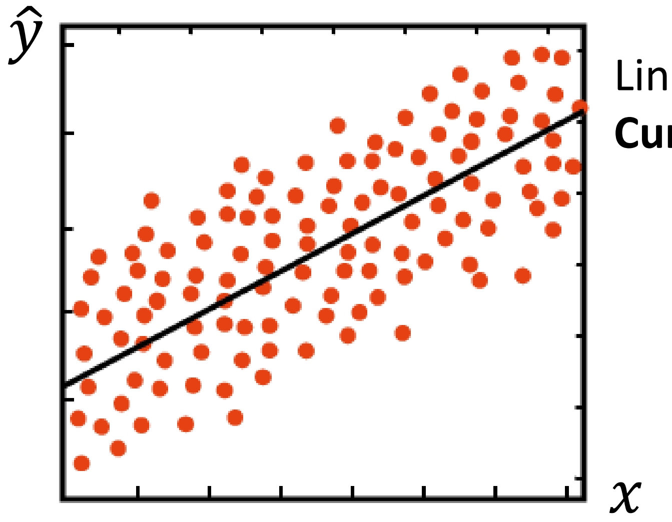
$\hat{y} \in \{y_A, y_B, \ldots\}$ multi-class

Line is a **Decision Boundary**

$$\hat{y} = M(\boldsymbol{x}), \quad \boldsymbol{x} = \vec{x} = (x_1, x_2)^\top$$

Regression

$\hat{y} \in \mathbb{R}$ any real number



Line is a **Curve Fit**

$$\hat{y} = M(x), \quad x = x_1$$

$$\hat{y} = M(\boldsymbol{x}), \quad \boldsymbol{x} = (x_1, \ldots, x_n)^\top$$

Image credit:
https://www.javatpoint.com/regression-vs-classification-in-machine-learning

Week 07b: Lecture Outline

1. Recap: Classification vs. Regression
2. Classification Metrics
3. Regression Metrics
    1. 1D regression: MSE, MAE
    2. Vector regression: Euclidean distance, Angular distance / Cosine Similarity
    3. Complex metrics for unstructured data

Note: intuition is opposite to "correctness".
- Longer distance means worse performance
- Smaller distance is better performance

Classification Correctness

Classification is correct when prediction $\hat{y}$ is the same as actual label y, i.e.,

$$\text{Correct} = [\hat{y} = y]$$

where

- $\hat{y} = M(\boldsymbol{x})$ is the predicted value from model M instance $\boldsymbol{x}$
- y is the ground truth value
- $[P] = \begin{cases} 1 & \text{if } P \text{ is true} \\ 0 & \text{otherwise} \end{cases}$ is the Iverson bracket notation for if/else

Regression Incorrectness “Difference”

Regression is less correct when prediction $\hat{y}$ is more different from actual label y, i.e.,

$$\text{Difference} = d(\hat{y}, y)$$

where

- $\hat{y} = M(\boldsymbol{x})$ is the predicted value from model M instance $\boldsymbol{x}$
- y is the ground truth value
- $d(\hat{y}, y)$ is some distance calculated between $\hat{y}$ and y

Difference metrics **per instance**

1. Simple difference:

$$d = \hat{y} - y$$

   But should we treat negative and positive difference differently? No.

2. Unsigned difference

   Absolute Error

$$d = |\hat{y} - y|$$

   Squared Error

$$d = (\hat{y} - y)^2$$

Squared Error penalizes larger differences **more** than Absolute Error

Average difference metrics **for test dataset**

Mean Absolute Error (MAE)

$$MAE = \frac{1}{m}\sum_{j=1}^{m}\left|\hat{y}_j - y_j\right|$$

Mean Squared Error (MSE)

$$MSE = \frac{1}{m}\sum_{j=1}^{m}\left(\hat{y}_j - y_j\right)^2$$

Root Mean Squared Error (RMSE)

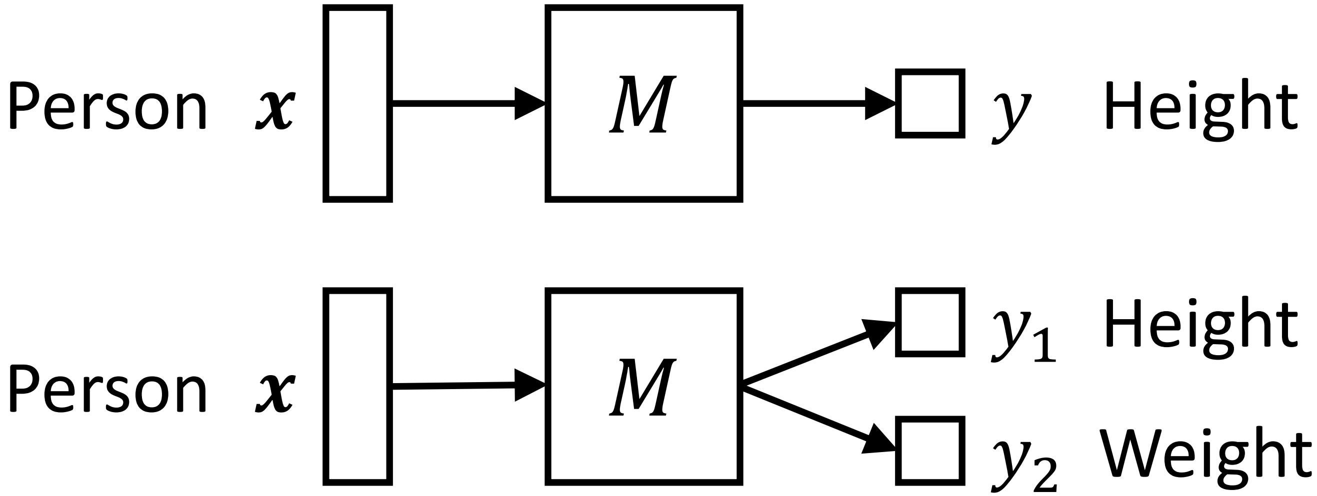
$$RMSE = \frac{1}{m}\sqrt{\sum_{j=1}^{m}\left(\hat{y}_j - y_j\right)^2}$$

MSE and RMSE penalize larger differences **more** than MAE

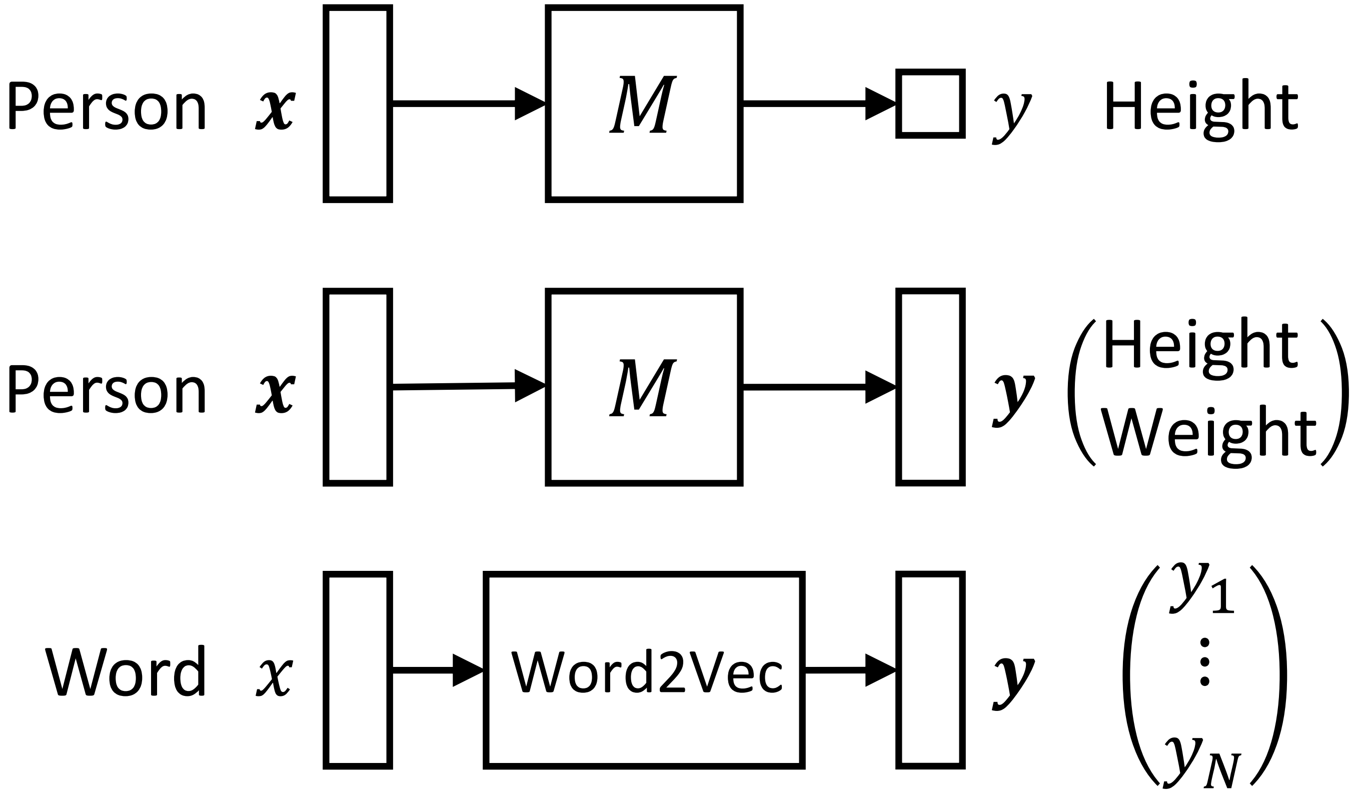
Week 07b: Lecture Outline

1. Recap: Classification vs. Regression
2. Classification Metrics
3. Regression Metrics
    1. 1D regression: MSE, MAE
    2. Vector regression: Euclidean distance, Angular distance / Cosine Similarity
    3. Complex metrics for unstructured data

Multi-task prediction



Multi-task prediction: predicting a vector $\boldsymbol{y}$

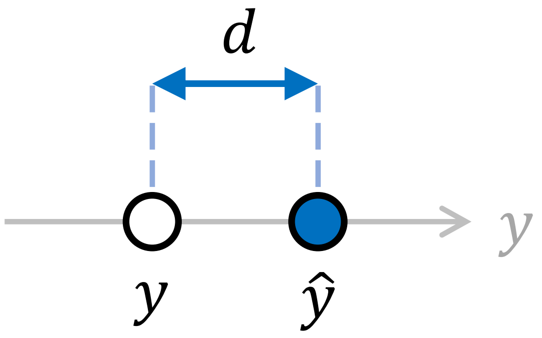


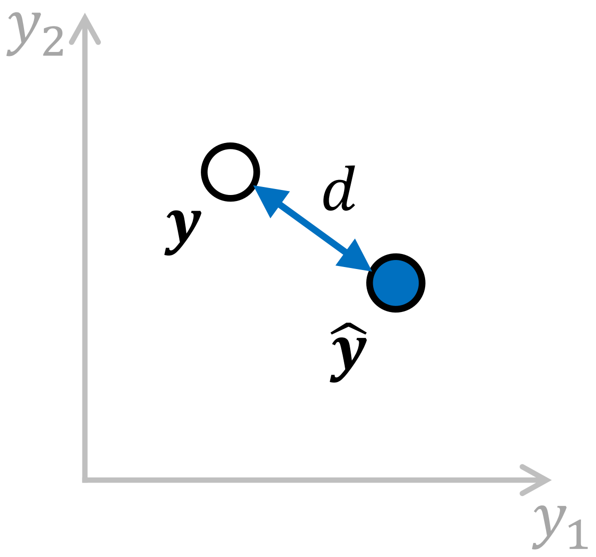
https://www.tensorflow.org/tutorials/text/word2vec

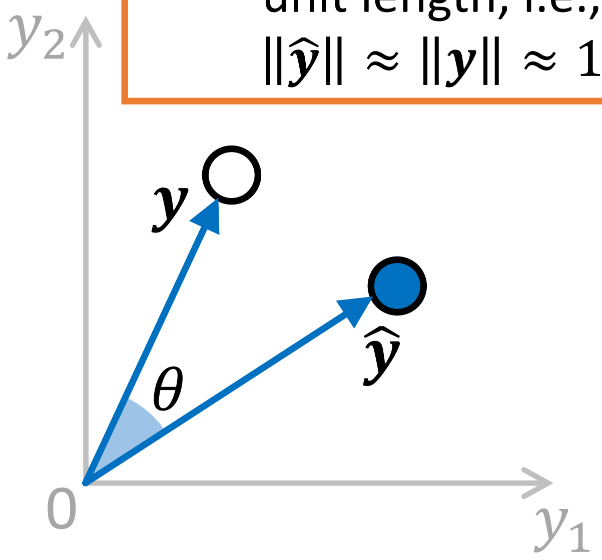
$\boldsymbol{y}$ is a vector representation of the text. Also known as "**embedding**"

Vector Distances and Similarity

Cosine similarity is often used for text embeddings, since their vectors are unit length, i.e.,
$\|\hat{\boldsymbol{y}}\| \approx \|\boldsymbol{y}\| \approx 1$







Squared Distance

$$d = (\hat{y} - y)^2$$

Euclidean Distance

$$d = \sqrt{(\hat{\boldsymbol{y}} - \boldsymbol{y})^\top (\hat{\boldsymbol{y}} - \boldsymbol{y})}$$

Dot Product

Cosine Similarity

$$s = \cos(\theta) = \frac{\hat{\boldsymbol{y}} \cdot \boldsymbol{y}}{\|\hat{\boldsymbol{y}}\| \|\boldsymbol{y}\|}$$

Angular Distance

$$\theta = \cos^{-1}(s)$$

Advanced Evaluation Metrics for Images, Time Series, Unstructured Data (with Deep Learning)

1. Similarity between (probability) distributions
    1. Kullback-Leibler Divergence
    2. Jensen-Shannon Distance
2. Similarity between images
    1. Mean Squared Error
    2. Peak Signal-to-Noise Ratio (PSNR)
    3. Structural Similarity Index Measure (SSIM)
    4. Pearson Correlation Coefficient
3. Segmentation (region) overlap
    1. Jaccard Index / Intersection-over-Union (IoU)

Won't be in the exam!

What did we learn for Evaluation?

1. Classification vs. Regression
2. Classification Metrics
    1. Accuracy
    2. Confusion Matrix, TP, TN, FP, FN
    3. Precision, Recall, F_1
    4. ROC, AUC
    5. Micro- and Macro-Averaging
    6. PR-AUC (Average Precision)
3. Regression Metrics
    1. 1D regression: MSE, MAE
    2. Vector regression: Euclidean distance, Angular distance / Cosine Similarity

Appropriate **evaluation metric** depends on prediction task and data issues.

Questions!