# Machine Learning

Pawel Wocjan

University of Central Florida

Fall 2020

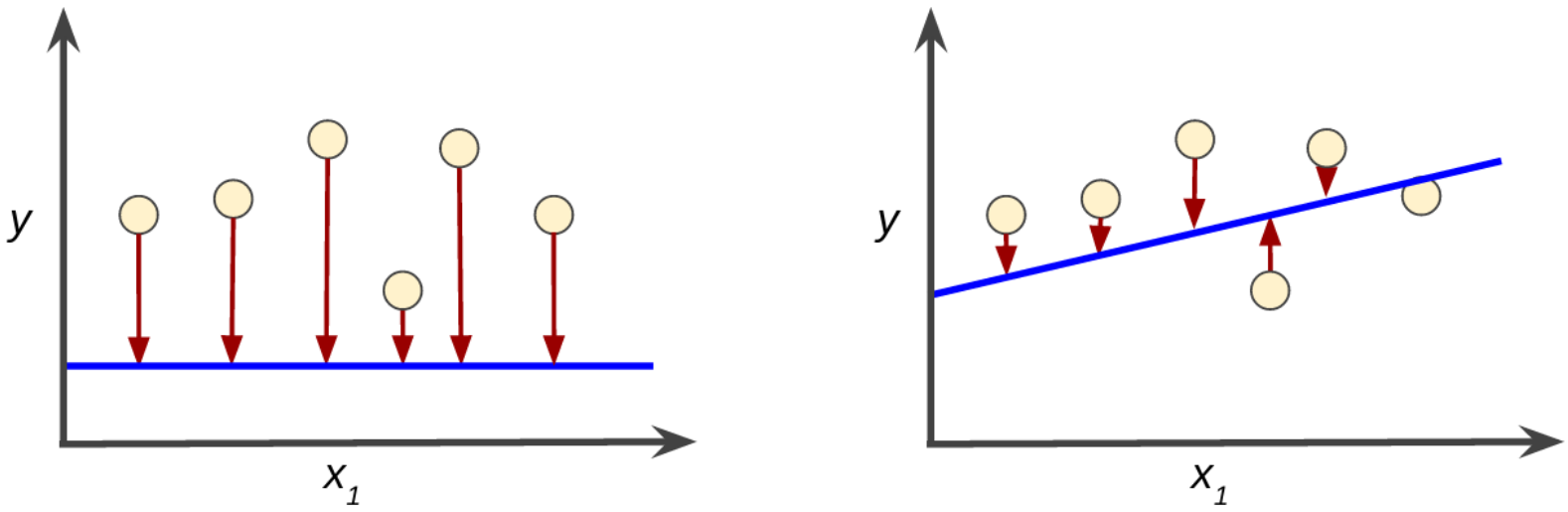
# Training and Loss

- **Training** a model simply means determining good values for all the weights and the bias that minimize the loss by examining many examples.
- **Loss** is the penalty for a bad prediction, that it, **loss** is a number indicating how bad the model's prediction was on a single example.

  If the model's prediction is perfect, the loss is zero; otherwise, the loss is greater.
- The goal of training is to find a set of weights and biases that have low loss, on average, across all examples.
- This process is called **empirical risk minimization**.

# Training and Loss

- For example, the figure below shows a high loss model on the left and a low loss model on the right.
  - The red arrows represent loss.
  - The blue lines represent predictions.



# Training and Loss

- Notice that the arrows in the left plot are much longer than their counterparts in the right plot.
- Clearly, the line in the right plot is a much better predictive model than the line in the left plot.
- The linear regression models we examine here use a loss function called **squared loss** (also known as $L_2$ loss).

# Training and Loss

- The squared loss for a single example the difference between the label (observation) $y$ and the prediction $\hat{y}$:

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

# Training and Loss

- **Mean square error (MSE)** is the average squared loss per example over the whole dataset

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

- $m$ is the number of examples.
- $x^{(i)} = (x_1^{(i)}, \ldots, x_n^{(i)})$ and $y^{(i)}$ are the features and the label of the $i$th example.
- $\hat{y}^{(i)}$ is the prediction of the model. More formally,

$$\hat{y}^{(i)} = f_\theta(x^{(i)})$$

where $\theta = (w_1, \ldots, w_n, b)$ are the parameters (weights and bias) of the model.

# Training and Loss

- Although MSE is commonly-used in machine learning, it is neither the only practical loss function nor the best loss function for all circumstances.

# Key Terms

- empirical risk minimization
- loss
- mean squared error
- squared loss
- training