Machine Learning

Pawel Wocjan

University of Central Florida

Fall 2020

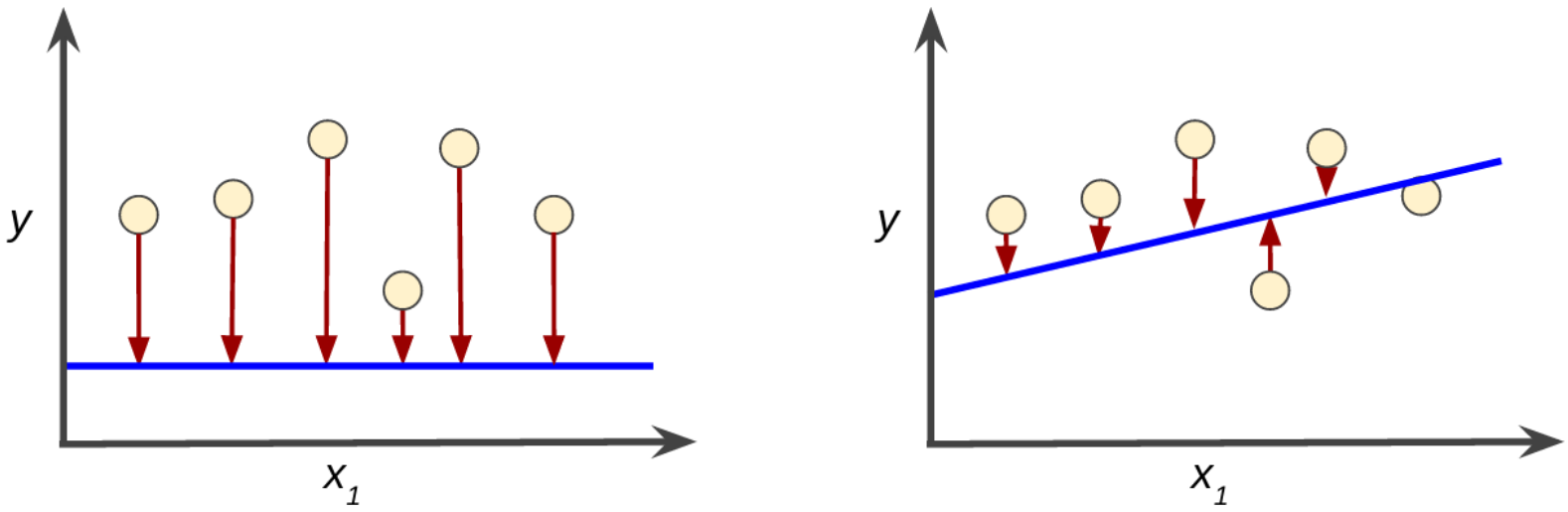
Training and Loss

- **Training** a model means examining the examples and adjusting the weights and the bias so that the loss is minimized.
- **Loss** is the penalty for a bad prediction, that it, it quantifies how bad the model's prediction was on a single example.

  If the prediction is perfect, the loss is zero; otherwise, it is greater than zero.
- 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 right blue line is a much better predictive model than the left blue line.
- The linear regression models we examine here use a loss function called **squared loss** (also known as L_2 loss).

Training and Loss

- Let $w = (b, w_1, \ldots, w_n)$ the parameters of the model (its weights and bias).
- Consider a labeled example with features $x = (x_1, \ldots, x_n)$ and label y.
- The model predicts

$$\hat{y} = f_w(x) = b + \sum_{j=1}^{n} w_j x_j$$

- 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

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

- 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 ith example.
- $\hat{y}^{(i)}$ is the prediction of the model. More formally,

$$\hat{y} = f_w(x^{(i)}) = b + \sum_{j=1}^{n} w_j x_j^{(i)}$$

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