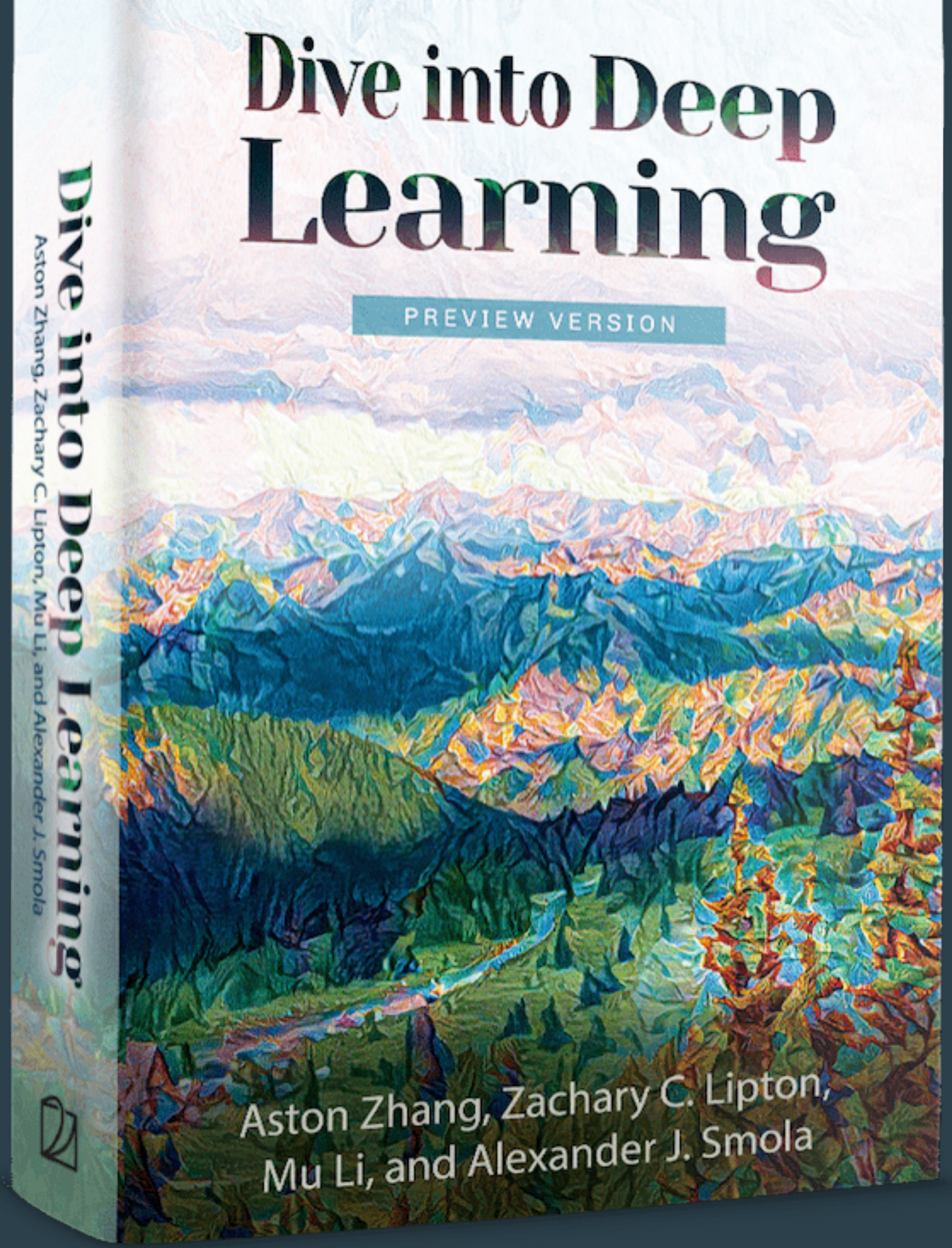
BLOCKCHAIN+AI

DeepLearning


Dive into Deep Learning
PREVIEW VERSION
Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola
Dive into Deep Learning
Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola

Chapter2

Linear Regression

- Regression: relationship between one or more independent variable with Dependent variable
- ML: often concerned with *prediction*.
- examples include predicting prices

Basic Elements

- *raining dataset* or *training set*
- *Label*
- *Features*

$$\mathbf{x}^{(i)} = [x_1^{(i)}, x_2^{(i)}]^\top$$

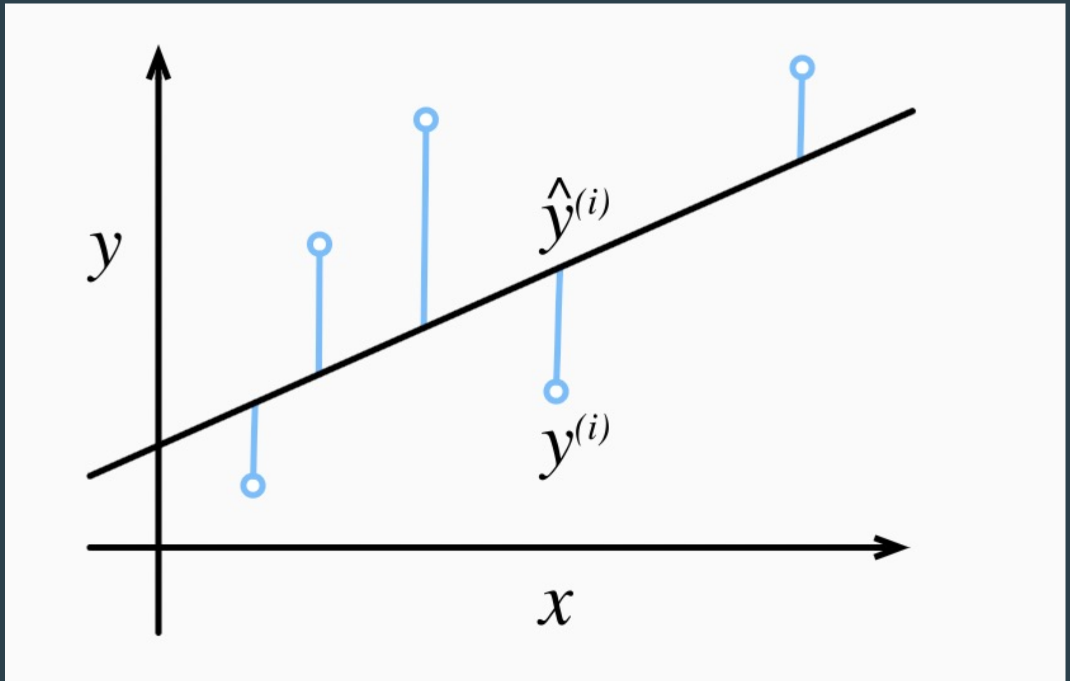
Linear Model

- Weights and Bias
- $\text{price} = w_{\text{area}} \cdot \text{area} + w_{\text{age}} \cdot \text{age} + b._{\text{price}}$
- output prediction is determined by the affine transformation of input features are *linear models*

$$\hat{y} = \mathbf{w}^\top \mathbf{x} + b.$$

- linear regression is to find the weight vector $\mathbf{w}$

Loss Function



Minibatch Stochastic Gradient Descent

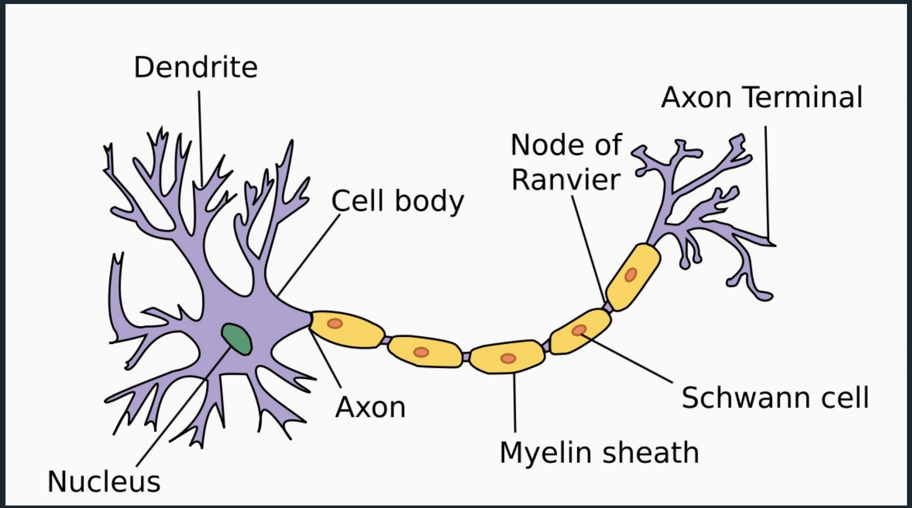
- iteratively reducing the error by updating the parameters in the direction that incrementally lowers the loss
- gradient descent consists of taking the derivative of the loss function,
- this can be extremely slow
- *minibatch stochastic gradient descent*

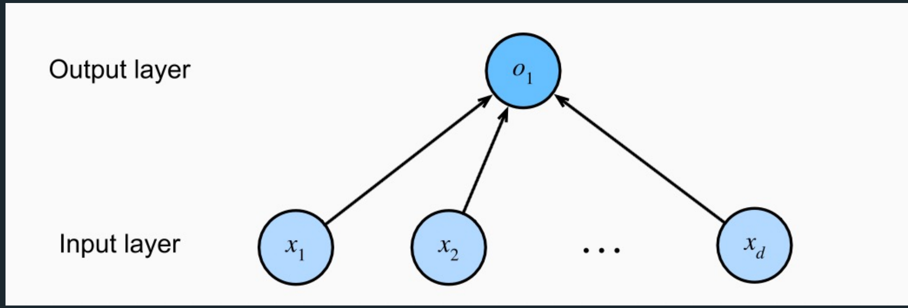
$$(\mathbf{w}, b) \leftarrow (\mathbf{w}, b) - \frac{\eta}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \partial_{(\mathbf{w}, b)} l^{(i)}(\mathbf{w}, b).$$

Vectorization for Speed

- **for** i **in** range(n):
    - c[i] = a[i] + b[i]
- reloaded + operator to compute the elementwise sum
- d = a + b
- Vectorizing code often yields order-of-magnitude speedup

From Linear Regression to Deep Networks





Summary

Key ingredients in a machine learning model are training data, a loss function, an optimization algorithm, and quite obviously, the model itself.

Vectorizing makes everything better (mostly math) and faster (mostly code).

Minimizing an objective function and performing maximum likelihood estimation can mean the same thing.

Linear regression models are neural networks, too.