Module 3

Linear and Logistic Regression with TensorFlow

Required

- Ramsundar, B., & Zadeh, R. B. (2018). *TensorFlow for deep learning: From linear regression to reinforcement learning*. (pp. Chapter 3). O'Reilly Media, Inc.
- Zhang, J., Jiang, J., & Liu, Y. (2018). Tensor learning and automated rank selection for regression-based video classification (Links to an external site.). *Multimedia Tools & Applications, 77*(22), 29213-29230.

Recommended

- Chapters 9 & 10 from *Deep Learning Pipeline: Building a Deep Learning Model with TensorFlow*
- Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction (Links to an external site.). *Computers & Education, 143*. https://doi.org/10.1016/j.compedu.2019.103676
- Yudistira, N., & Kurita, T. (2018). Correlation Net: Spatiotemporal multimodal deep learning for action recognition (Links to an external site.). https://doi.org/10.1016/j.image.2019.115731
- LinkedIn Learning (Director). Building Deep Learning Applications with Keras 2.0 (Links to an external site.) [Video]. LinkedInLearning.

▼ TensorFlow is a programming language based on what math?

vector calculus and statistics, namely logistic regression

▼ What is convergence in deep learning?

the process of achieving a prediction or determination from a dataset that is improving on a less-than-perfect prediction, which is derived from a small subset of given information

▼ What is the purpose of convergence?

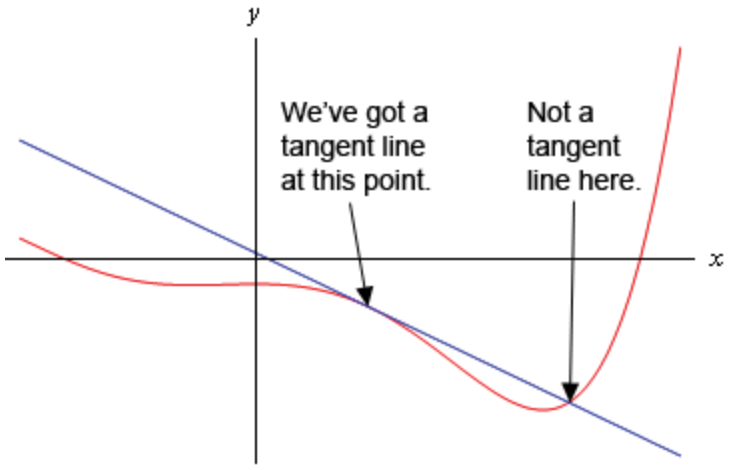
TensorFlow implements the convergence approach to minimize a "loss function," which is essentially the "difference" between the actual answer and the current estimated answer

▼ What is the technique of minimizing the loss function?

gradient descent

▼ What is a continuous function?

- one for which a tangent line to any point on that function has a finite slope, such as *f(x)* = tan(x) on the interval [−π3,π3]
- the tangent line (or simply tangent) to a plane curve at a given point is the straight line that "just touches" the curve at that point



▼ What is function differentiability?

- the plot of the function does not have sharp corners or turns
- an example would be f(x) = x2 (a parabola)

▼ What is the *derivative* of differentiable function?

it is another function that provides the slope of the tangent line to any point on the function *f*

▼ TensorFlow is the implementation of an optimization algorithm that does what?

- it computes a series of points that track along the graph of a continuous function, which leads toward the (local) minima of *f*
- at the exact minima point, a continuous function will have derivative zero

▼ What is the gradient of f (denoted ∇f)?

- it is the multi-dimensional analog for the derivative of a function of a single variable
- machine learning is a process of computing the minimum point of a function, regardless of how many variables the function may have
- the key is to use the gradient function for multi-variable functions

▼ Why are loss functions central to machine learning?

- To solve a machine learning problem, a data scientist must find a way of building a function whose minima translates to a useful solution for an actual machine learning problem.
- Machine learning design entails thoughtful creation of a “good” *loss function* to solve a specified problem.

▼ What is the loss function based on?

- The loss function ***L*** is based on datapoints x and labels y and is formulated as ***L*** *(x,y)*.
- In this formula, **x and y are tensors, and the loss function maps those tensors to a scalar.**
- If (xi,yi) are data available at *i* and there are *N* points, then

$$L(x, y) \ = \sum_{i=1}^{N} \ L_i(x_i, \ y_i)$$

and the gradient of the loss function is:

$$\nabla L(x, y) \ = \sum_{i=1}^{N} \ \nabla L_i(x_i, \ y_i)$$

▼ A loss function for a machine learning system is what?

the sum of individual loss functions at different data points

▼ ML algorithms can be broadly categorized into what?

- supervised or unsupervised
- supervised problems have both datapoints x and labels y available, while unsupervised problems have only datapoints x without labels y

▼ Supervised ML can be broken up into what subproblems?

classification and regression

▼ What is a classification problem?

A classification problem is one where you want a machine learning system that assigns a discrete label, say 0/1 (or more generally 0,…,n) to a datapoint.

▼ What is a regression problem?

A type of classification problem is *regression*, where you assign a real valued label from R to a given datapoint.

▼ What is L^2 loss function used for?

regression problems

▼ What is the L^2 loss function?

- It is defined as the magnitude of a vector
- The L^2 function where **a** is a vector of length *N:*

$$\|a\|_2 = \sqrt{\sum_{i=1}^{N} a_i^2}$$

- Note: This norm is also used to define the distance between two vectors:

$$\|a - b\|_2 = \sum_{i=1}^{N} (a_i - b_i)_2$$

- Applying L^2 as a distance measurement is useful for solving regression problems in supervised machine learning. Let us suppose that *x* is a collection of data and *y* the associated labels. Let *f* be some differentiable function that realizes a machine learning model. Then, to use *f* to predict *y*, construct the L^2 loss function:

$$L(x, y) = \|f(x) - y\|_2$$

- Key:
  - The idea is to use the differentiability of *L(x,y)*
    to find the point that minimizes it. This point is the machine learning result.
  - The same results can be attained if the loss function is squared, which makes for easier computations.
  - This technique works for simple machine learning models with few parameters.