BREAK THROUGH TECH

Machine Learning Foundations

Lab 4

Week of June 16

Today's Agenda

(15 minutes)	Icebreaker
(30 minutes)	Week 4 Concept Overview + Q&A
(20 minutes)	Breakout Groups: Big Picture Questions
(10 minutes)	Class Discussion
(10 minutes)	Break
(35 minutes)	Breakout Groups: Lab Assignment Working
Session 1	
(5 minutes)	Working Session 1 Debrief
(35 minutes)	Breakout Groups: Lab Assignment Working
Session 2	
(5 minutes)	Working Session 2 Debrief
(5 minutes)	Concluding Remarks & Survey

BREAK THROUGH TECH

Guess the Prompt Game

Instructions

Objective: Your goal is to predict the prompt based on the given image that was generated by DALL-E

Analyze the features and make a prediction about the outcome.

Make your predictions. Consider the significance of each feature in the image.

- Share your thoughts, insights, and reasoning in the chat box
- We'll reveal the correct outcome for the slide..

A penguin breakdancing to country music in a jungle

A group of lion nerds farming


Pandas Python Library
Advanced AI

A panda creating top of the line AI using the pandas python library

Challenge Mode!

Two animals doing an activity together, make it crazy and random

Send a prompt via private message to me on zoom! I will pick one to generate and have us guess.

a determined goldfish rides a bike to the aquarium to save his family

BREAK THROUGH TECH

Week 4 Concept Overview + Q&A

Parametric vs Non-parametric models

- DT and kNN: Non-parametric
- Parametric model: Model is represented by an equation
- Simple example: $y = ax + b$
- Parameters: a, b
- Logistic regression, support vector machines, neural networks
- GPT 4: Approximately 1,760,000,000,000 parameters
- Loss function = Function that represents error rate of model
- Loss function determines how to update parameters

Logistic regression: Prediction

Problem: Features (x_1, x_2); Predict T/F label

[Obtained by learning]

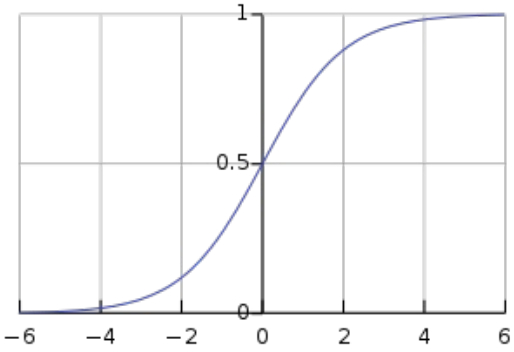
Parameters/weights

w_1	w_2	w_0
0.5	2.1	1.5

Test instances

x_1	x_2	Label T/F
23.5	-7.2	?
-1	1.5	?

1. Linear step: $L = w_1 * x_1 + w_2 * x_2 + w_0$ -1.87 4.15
2. Sigmoid squash: $p = \frac{1}{1+e^{-L}}$ 0.13 0.97
3. Threshold: $p \geq 0.5$? F T



First instance:
0.5*23.5 + 2.1*(-7.2) + 1.5 = -1.87

Second instance:
0.5*(-1) + 2.1*(1.5) + 1.5 = 4.15

Learning weights: Loss Functions

1. Log loss

$$L_{\mathrm{LL}} = -\frac{1}{N}\sum_{i=1}^{N}\Big(y_i \log(P_i) + (1-y_i)\log(1-P_i)\Big)$$

2. Mean Squared Error

$$L_{\mathrm{MSE}} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

3. Zero-One Loss

$$L_{0/1} = \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}(\hat{y}_i \neq y_i) \qquad \mathbb{I}(\hat{y}_i \neq y_i) = \begin{cases} 1 & \text{if } \hat{y}_i \neq y_i \\ 0 & \text{if } \hat{y}_i = y_i \end{cases}$$

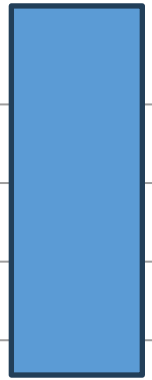
Loss Functions: Updating weights

Consider single variable x_1 and its weight w_1

Initial (random) value of w_1: 3.15

Training set

x_1		Label 1 or 0
1.89		1
-2.1		0
3.05		0



$p_1 = 0.18$ Increase or decrease w_1?

$p_1 = 0.24$ Increase or decrease w_1?

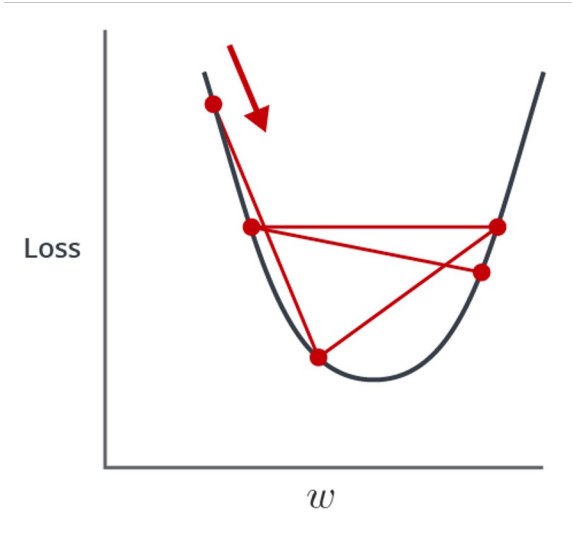
$p_1 = 0.87$ Increase or decrease w_1?

Update:

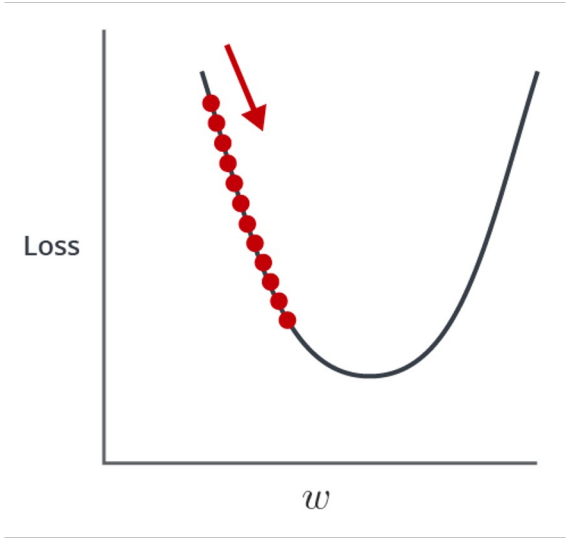
$$w_1 = w_1 + \gamma(y_i - p_i)x_i$$

If the linear step is more positive, then final probability closer to 1
If the linear step is more negative, then final probability closer to 0

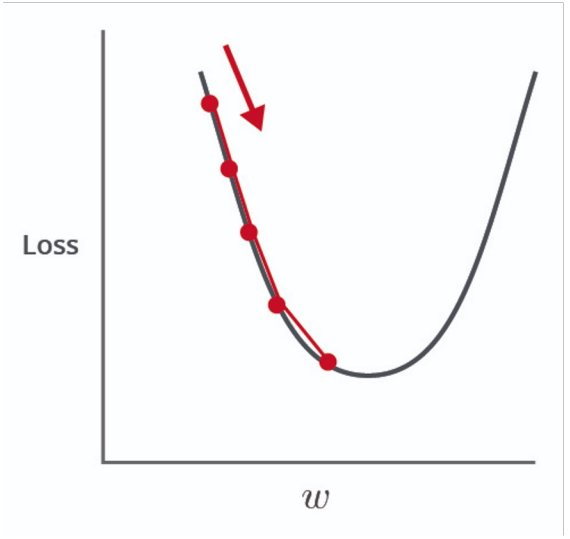
Learning Rate γ



Too high



Too low

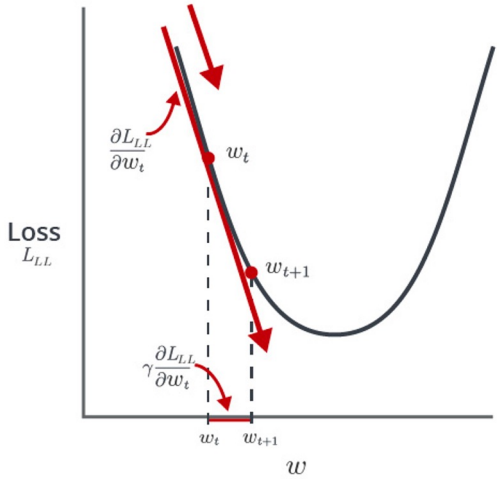


Ideal

Determining Learning Rate γ

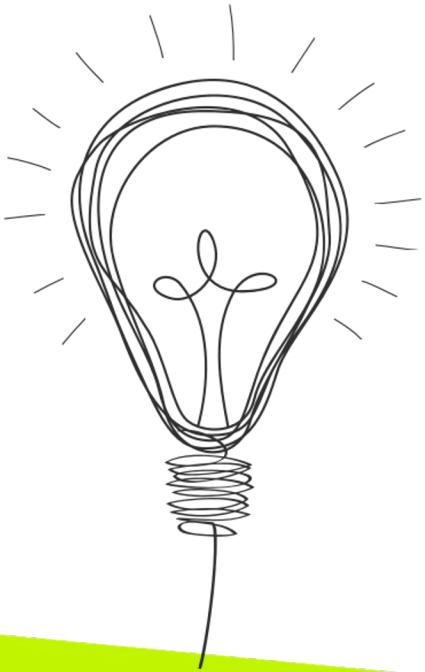
Use the Hessian function to compute optimal

- Second derivative of function, represents curvature
- Steep curve → small rate, gradual curve, → larger rate



Questions & Answers

What questions do you have about the online content this week?



BREAK THROUGH TECH

Breakout Groups: Big Picture Questions

Big Picture Questions

You have 20 minutes to discuss the following questions within your breakout groups:

- How do decision trees and logistic regression differ in terms of their underlying algorithms or models?
- What are the differences in the type of output or predictions generated by decision trees and logistic regression?
- In which scenarios would logistic regression be more suitable or preferable compared to a decision tree?
- Why is it good or useful to estimate the probability of something occurring, as opposed to just making a binary classification decision?

BREAK THROUGH TECH

Class Discussion

BREAK
THROUGH
TECH

Break

BREAK THROUGH TECH

Breakout Groups: Lab Assignment

Lab 4

In this lab, you will:

- Write a Python class that will train a logistic regression model and make predictions
- Build your DataFrame and define your ML problem
- Create labeled examples from the data set
- Train a logistic regression model using your class
- Compare your Python implementation to scikit-learn's implementation.

Lab 4

Lab 4: ML Life Cycle: Modeling

Building a Logistic Regression Model From Scratch

In [ ]:
```
import pandas as pd
import numpy as np
import os
from sklearn.linear_model import LogisticRegression
```

In this lab, you will continue working with the modeling phase of the machine learning life cycle. You will take what you have learned about gradient descent and write a Python class from scratch to train a logistic regression model. You will implement the various mathematical functions learned in the course, such as the gradient and Hessian of the log loss.

In the course videos, we presented functions that compute the log loss gradient and Hessian and that implement gradient descent for logistic regression. You will do similar work here, only we'll refactor the code to improve its generality.

You will complete the following tasks:

1. Build a class that can:
   - fit a logistic regression model given training data
   - make predictions
2. Build your DataFrame and define your ML problem:
   - Load the Airbnb "listings" data set into a DataFrame
   - Define the label - what are you are predicting?
   - Identify features
3. Create labeled examples from the data set
4. Train a logistic regression classifier using your class
5. Benchmark our class against scikit-learn's logistic regression class

BREAK THROUGH TECH

Working Session 1 Debrief

Lab Debrief

So far,

- What did you enjoy about this lab?
- What did you find difficult about this lab?
- What questions do you still have about this lab?

BREAK THROUGH TECH

Working Session 2 Debrief

Lab Debrief

- What did you enjoy about this lab?
- What did you find hard about this lab?
- What questions do you still have about this lab?
- How did you approach problem-solving during the exercise?
- What would you do differently if you were to repeat the exercise?

BREAK
THROUGH
TECH

Concluding Remarks

Concluding Remarks

- Key takeaways
- Additional resources

Next week

In the following week, you will:

- Understand the importance of model selection in machine learning
- Choose model evaluation metrics that are appropriate for the application
- Choose appropriate model candidates and hyperparameters for testing
- Set up training/validation/test splits for model selection
- Apply feature selection techniques to get a better-performing model
- Explore how to deploy, host, and monitor your model

And in the lab, you will:

- Build your DataFrame and define your ML problem
- Create labeled examples from the data set, and split the data into training and test data sets
- Train, test and evaluate a logistic regression model using scikit-learn's default hyperparameter value for C.
- Find the optimal logistic regression model using GridSearchCV.
- Train, test and evaluate the optimal logistic regression model.
- Plot the precision-recall curve and the ROC, then compute the AUC for both models.
- Practice the SelectKBest feature selection method.
- Save your best performing model to a PKL file, and add the model and dataset to your GitHub repository.

BREAK THROUGH TECH

Content + Lab Feedback Survey

Content + Lab Feedback Survey

To complete your lab, please answer the following questions about BOTH your online modules and your lab experience. Your input will help pay it forward to the Break Through Tech student community by enabling us to continuously improve the learning experience that we provide to our community.

Thank you for your thoughtful feedback!

https://forms.gle/eUQQZgS6BPRpqgZ7A