Lecture 22: Communication and Conclusion

Firas Moosvi

Clicker (Survival analysis recap)

Select all of the following statements which are TRUE.

- a. Right censoring occurs when the endpoint of event has not been observed for all study subjects by the end of the study period.
- b. Right censoring implies that the data is missing completely at random.
- c. In the presence of right-censored data, binary classification models can be applied directly without any modifications or special considerations.
- d. If we apply the `Ridge` regression model to predict tenure in right censored data, we are likely to underestimate it because the tenure observed in our data is shorter than what it would be in reality.

Recap

- What is right-censored data?
- What happens when we treat right-censored data the same as “regular” data?
  - Predicting churn vs. no churn
  - Predicting tenure
    - Throw away people who haven’t churned
    - Assume everyone churns today
- Survival analysis encompasses predicting both churn and tenure and deals with censoring and can make rich and interesting predictions!
  - We can get survival curves which show the probability of survival over time.
  - KM model ⟶ doesn’t look at features
  - CPH model ⟶ like linear regression, does look at the features and provides coefficients associated with each feature

Why communication?

Why spend a whole lecture on this?

- **Great technical work often dies silently due to poor communication.**
- Most ML work happens in teams with diverse backgrounds.
- Decisions, budgets, and user trust depend on how you present results.
- Effective communication ⟶ adoption and impact

Is this misleading?

What additional information would you need to evaluate the validity of this claim?

Main issues in ML-related communication

- Overclaiming and unclear limitations
- No baseline or missing context for metrics
- Metric dumping without a narrative or decision link
- Jargon/abstraction mismatch to the audience
- Unexplained predictions; weak explainability path
- Hidden assumptions or data leakage
- Probability misinterpretation (e.g., `predict_proba=0.9` ≠ 90% certain)
- Unreliable test error due to weak validation

Scenario discussion: What happens if...

Pick one scenario: Discuss 2 negative consequences and 1 thing you'd do to prevent them.

1. You build an amazing model but fail to clearly communicate its value or results to your manager.
2. You present a 98% accuracy without mentioning the trivial baseline is 97.5%.
3. You say: "SHAP values show nonlinear feature interactions" to a non-technical stakeholder and stop there.
4. A user asks why they were denied a loan; you give no explanation of the model's decision.
5. You hide uncertainty and overpromise deployment success.

Principles of good communicaiton

Grid search activity

Go to this Google doc: https://tinyurl.com/5n8xf5yj

Explanation 1:
https://tinyurl.com/msk2cfkb

Machine learning algorithms, like an airplane's cockpit, typically involve a bunch of knobs and switches that need to be set.

Explanation 2:
https://tinyurl.com/mt2z9ey5

An introduction to Grid Search

Krishni Follow 2 min read · Jan 5, 2019

512 2

This article will explain in simple terms what grid search is and how to implement grid search using `sklearn` in python.

Discussion questions

- What do you like about each explanation?
- What do you dislike about each explanation?
- What do you think is the intended audience for each explanation?
- Which explanation do you think is more effective overall for someone on Day 1 of CPSC 330?
- Each explanation has an image. Which one is more effective? What are the pros/cons?
- Each explanation has some sample code. Which one is more effective? What are the pros/cons?

Concepts then labels, not the other way around

Explanation 1: Machine learning algorithms, like an airplane's cockpit, typically involve a bunch of knobs and switches that need to be set.

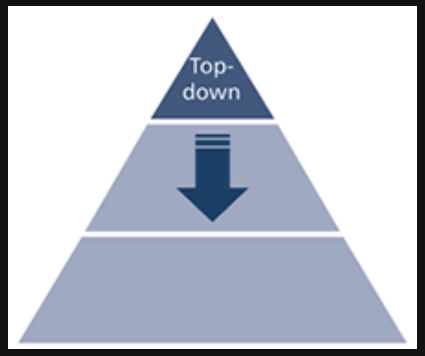
Explanation 2: Grid search is the process of performing hyper parameter tuning in order to determine the optimal values for a given model.

The effectiveness of these different statements depend on your audience.

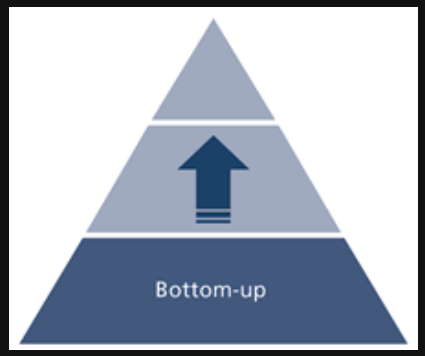
Concepts then labels, not the other way around

Top down vs. bottom up

- Start with the big picture
- Then gradually reveal the structure and key components



- Start with the details
- Build to the big picture



In the previous explanations, which one represented a bottom-up explanation and which one a top-down explanation?

New ideas in small chunks

The hidden structure in the first explanation

1. The concept of setting a bunch of values.
2. Random forest example.
3. The problem / pain point.
4. The solution.
5. How it works - high level.
6. How it works - written example.
7. How it works - code example.
8. The name of what we were discussing all this time.

Reuse running examples

Effective explanations often use the same example throughout the text and code. This helps readers follow the line of reasoning.

Approach from all angles

- When we're trying to draw mental boundaries around a concept, it's helpful to see examples on all sides of those boundaries
- It would have been nice to include
  - Performance with and without hyperparameter tuning.
  - Other types of hyperparameter tuning (e.g. `RandomizedSearchCV`).

When experimenting, show the results asap

The first explanation shows the output of the code, whereas the second does not. This is easy to do and makes a big difference.

It's not about you

- Interesting to you != useful to the reader (aka it's not about you)
- Examine the hidden intention of wanting to include something that's not important
  - Am I trying to sound smart or prove I know something?
  - Am I afraid that leaving it out makes the work look too simple?
  - Am I adding it because I spent time on it and want that effort to be visible?
  - Am I overexplaining because I'm worried the audience will judge me?

> **If it doesn't serve the audience, it's noise.**

Core questions you must be ready to answer

- What does this result mean (in plain language)?
- When does the model work? When does it fail? (failure modes)
- Why did it make this prediction? (explainability path)
- What are the risks & consequences of using it?
- How does it compare to doing nothing or current practice?
- What is the cost to maintain / retrain / monitor?

Quick checklist (use before presenting)

1. ▪ Who is my audience and what do they care about?
2. ▪ What decision do I want them to make?
3. ▪ What baseline(s) am I comparing against?
4. ▪ What caveats or limitations must I disclose?
5. ▪ What is the recommended next action?
6. ▪ How will we monitor after deployment?

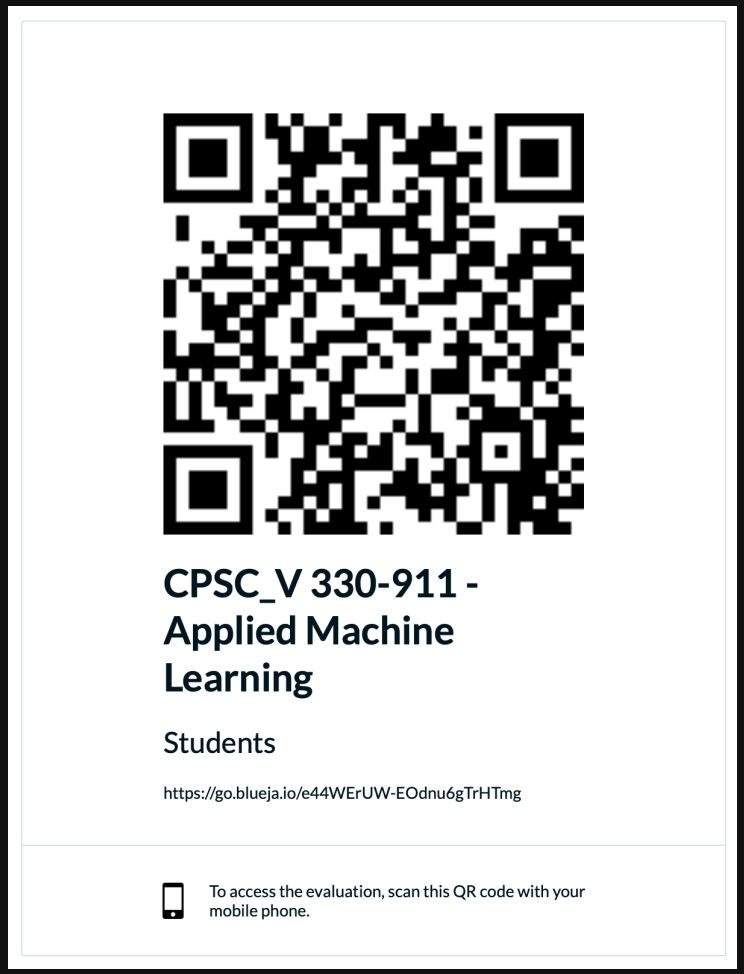
Poor vs. Effective communication

Which one is poor and which one is effective? Why?

Poor vs. Effective communication

Course evaluations (~10 mins)

Complete the SEoI for this course here



ML and decision making

? ? Questions for you

Imagine you are tasked with developing a recommender system for YouTube. You possess data on which users clicked on which videos. After spending considerable time building a recommender system using this data, you realize it isn't producing high-quality recommendations. What could be the reasons for this?

Think beyond the data that's given to you

Questions you have to consider:

- Who is the decision maker?
- What are their objectives?
- What are their alternatives?
- What is their context?
- What data do I need?

Decisions involve a few key pieces

- The **decision variable**: the variable that is manipulated through the decision.
  - E.g. how much should I sell my house for? (numeric)
  - E.g. should I sell my house? (categorical)
- The decision-maker's **objectives**: the variables that the decision-maker ultimately cares about, and wishes to manipulate indirectly through the decision variable.
  - E.g. my total profit, time to sale, etc.
- The **context**: the variables that mediate the relationship between the decision variable and the objectives.
  - E.g. the housing market, cost of marketing it, my timeline, etc.

Confidence and `predict_proba`

- What does it mean to be “confident” in your results?
- When you perform analysis, you are responsible for many judgment calls.
- Your results will be different than others.
- As you make these judgments and start to form conclusions, how can you recognize your own uncertainties about the data so that you can communicate confidently?

Let's imagine that the following claim is true:

> Vancouver has the highest cost of living of all cities in Canada.

Now let's consider a few beliefs we could hold:

1. Vancouver has the highest cost of living of all cities in Canada. **I am 95% sure of this.**
2. Vancouver has the highest cost of living of all cities in Canada. **I am 55% sure of this.**

The part is bold is called a credence. Which belief is better?

But what if it's actually Toronto that has the highest cost of living in Canada?

1. Vancouver has the highest cost of living of all cities in Canada. **I am 95% sure of this.**
2. Vancouver has the highest cost of living of all cities in Canada. **I am 55% sure of this.**

Which belief is better now?

We don't just want to be right. We want to be confident when we're right and hesitant when we're wrong.

In our final exam, imagine if, along with your answers, we ask you to also provide a confidence score for each. This would involve rating how sure you are about each answer, perhaps on a percentage scale from 0% (completely unsure) to 100% (completely sure). This method not only assesses your knowledge but also your awareness of your own understanding, potentially impacting the grading process and highlighting areas for improvement. Who supports this idea 😉?

Loss in machine learning

When you call `fit` for `LogisticRegression` it has similar preferences:

correct and confident
> correct and hesitant
> incorrect and hesitant
> incorrect and confident

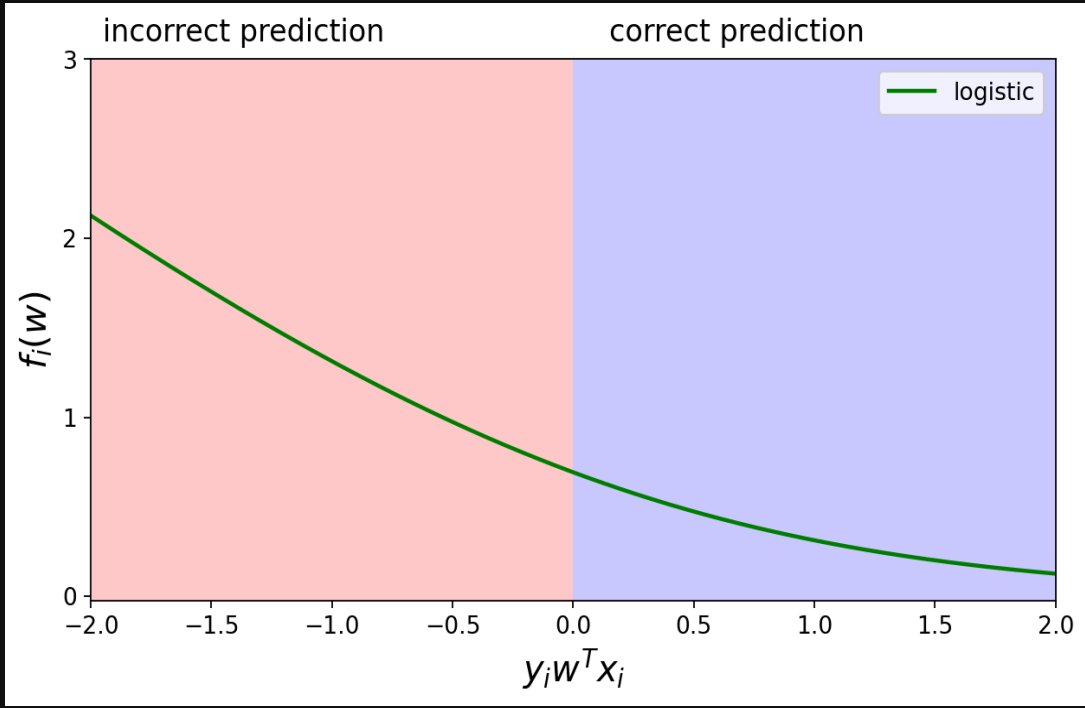
- This is a “loss” or “error” function like mean squared error, so lower values are better.
- When you call `fit` it tries to minimize this metric.

What should be the loss?

Consider the following made-up classification example where target (true `y`) is binary: -1 or 1. The true y (`y_true`) and models raw scores ($w^T x_i$) are given to you. You want to figure out how do you want to punish the mistakes made by the current model. How will you punish the model in each case?

	y_true	raw score (w^Tx_i)	correct? (yes/no)	confident/hesitant?	punishment
0	1	10.00	yes	confident	None
1	1	0.51	yes	hesitant	small punishment
2	1	-0.10	no	hesitant	
3	1	-10.00	no	confident	
4	-1	-12.00	yes	confident	
5	-1	-1.00	yes	hesistant	
6	-1	0.40	no	hesitant	
7	-1	18.00	no	confident	

Logistic regression loss



- confident and correct ⟶ smaller loss
- hesitant and correct ⟶ a bit higher loss
- hesitant and incorrect ⟶ even higher loss
- confident and incorrect ⟶ high loss

Misleding visualizations

This chart is attempting to suggest a relationship between childhood MMR vaccination rates and the prevalence of autism spectrum disorders (AD/ASD) across several countries.

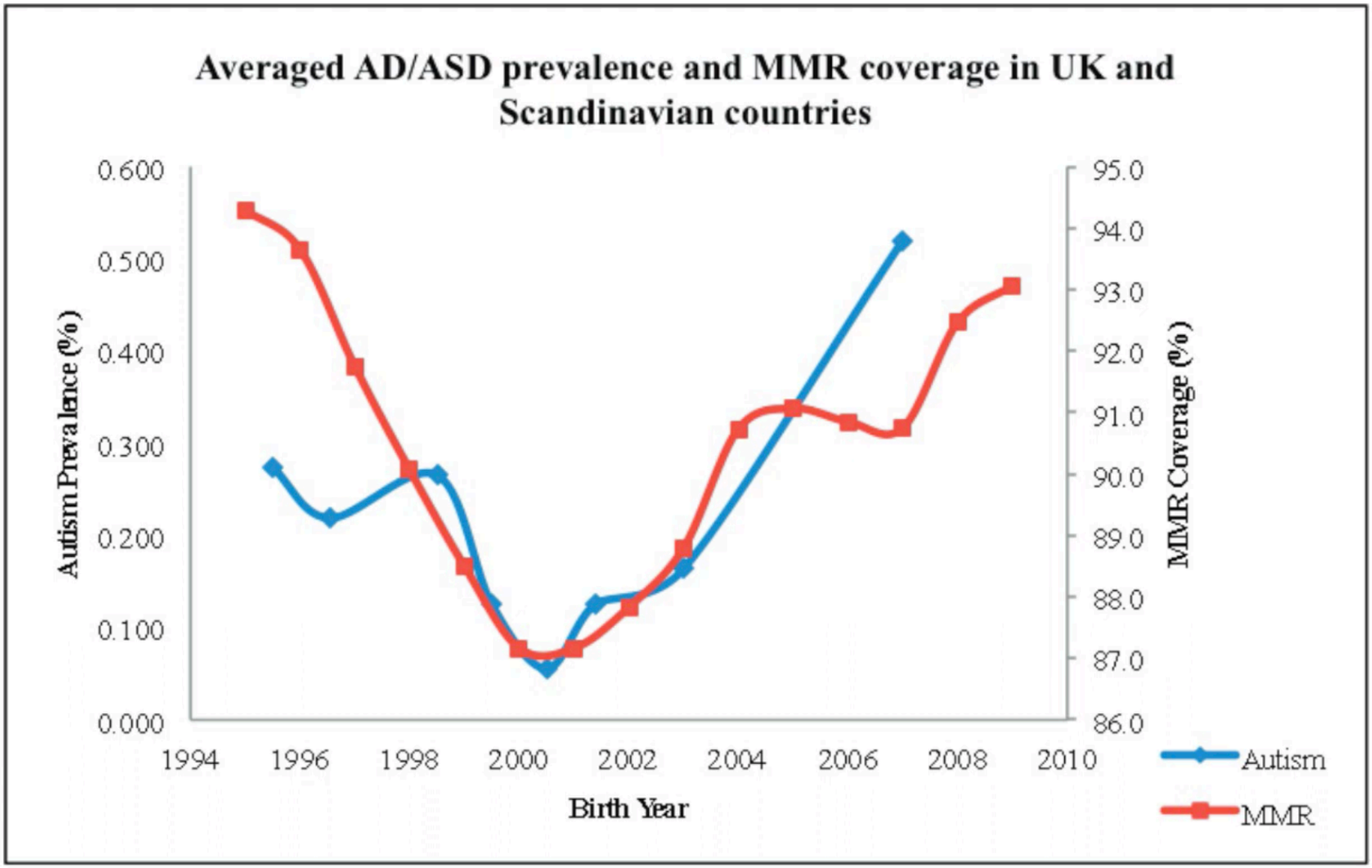


Figure 1-Averaged AD/ASD prevalence and MMR coverage in UK, Norway and Sweden. Both MMR and AD/ASD data are normalized to the maximum coverage/prevalence during the time period of this analysis.

Visualizing your data and results could be very powerful but at the same time can be misleading if not done properly.

Examples

Some examples from Calling BS visualization videos:

- Dataviz in the popular media: modern NYT
- Misleading axes: vaccines
- Manipulating bin sizes: tax dollars
- Dataviz ducks: drinking water
- Glass slippers: internet marketing tree
- The principle of proportional ink: most read books

Things to watch out for

- Chopping off the x-axis
  - the practice of starting the x-axis (or sometimes the y-axis) at a value other than zero to exaggerate the changes in the data
- Saturate the axes
  - where the axes are set to ranges that are too narrow or too wide for the data being presented making it difficult to identify patterns
- Bar chart for a cherry-picked values
- Different y-axes

What did we learn today?

Principles of effective communication

- Concepts then labels, not the other way around
- Bottom-up explanations
- New ideas in small chunks
- Reuse your running examples
- Approaches from all angles
- When experimenting, show the results asap
- **It's not about you.**
- Decision variables, objectives, and context.
- Expressing your confidence about the results
- Misleading visualizations.

? ? Questions for you

Imagine you've created a machine learning model and are eager to share it with others. Consider the following scenarios for sharing your model:

- **To a non-technical Audience:** How would you present your model to friends and family who may not have a technical background?
- **To a technical audience:** How would you share your model with peers or professionals in the field who have a technical understanding of machine learning?
- **In an academic or research setting:** How would you disseminate your model within academic or research communities?

Try out this moment predictor

https://cpsc330-moment-predictor.onrender.com/

- In this lecture, I will show you how to set up/develop this.

What is deployment?

- After we train a model, we want to use it!
- The user likely does not want to install your Python stack, train your model.
- You don't necessarily want to share their dataset.
- So we need to do two things:
  1. Save/store your model for later use.
  2. Make the saved model conveniently accessible.

We will use the tools below for

- Saving the model: We will use Joblib
- Making the saved model conveniently accessible: Flask & render

Full examples of deploying an app

For this demo, each student should click this link to create a new repo in their accounts, then clone that repo locally to follow the steps to deloy an app.

Web App

See notes fom class as well as the lecture demo.

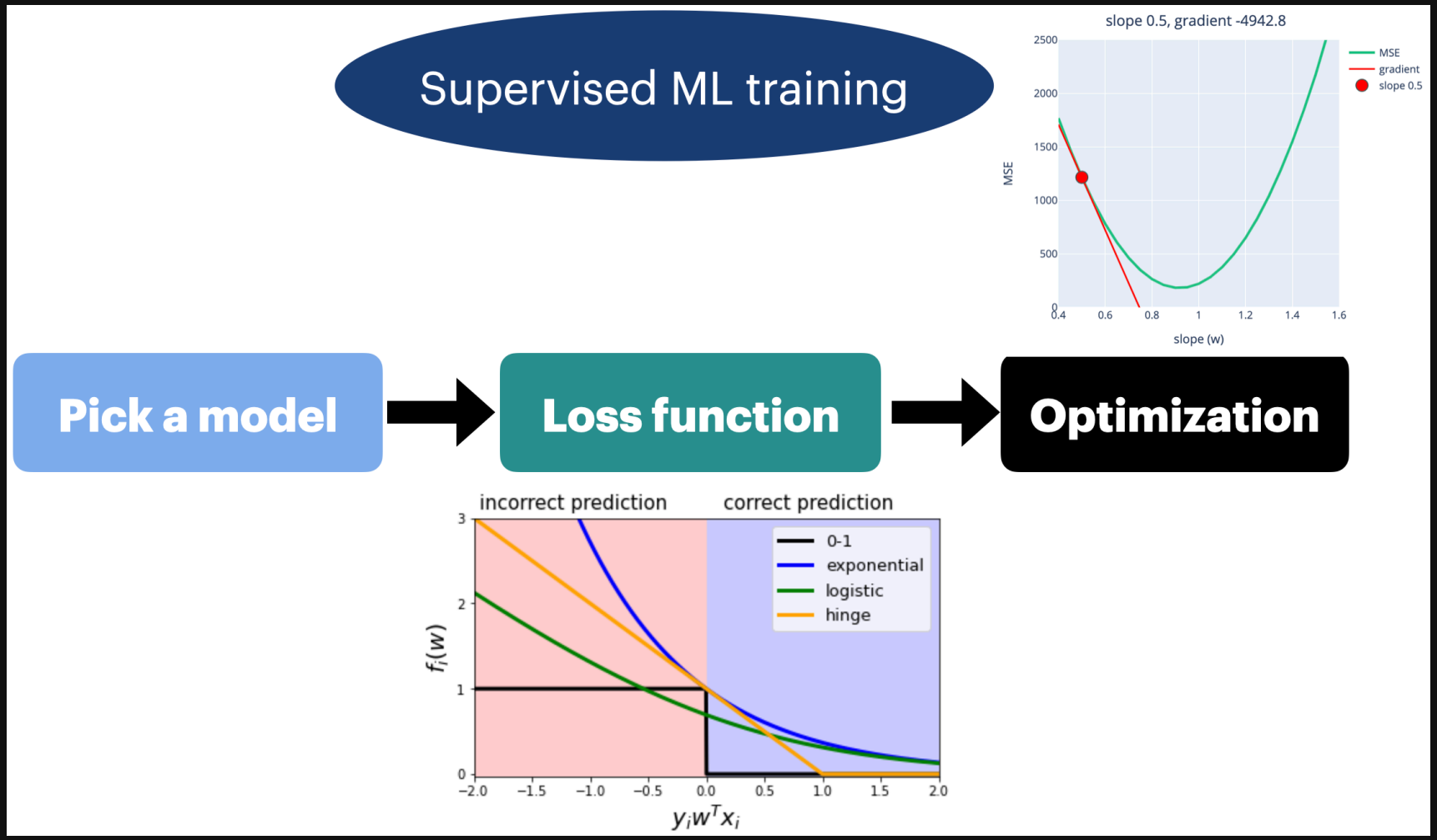
API

What did we cover

- Part 1: Supervised learning on tabular data: ML fundamentals, preprocessing and data encoding, a bunch of models, evaluation metrics, feature importances and model transparency, feature selection, hyperparameter optimization
- Part 2: Dealing with other non-tabular data types: Clustering, recommender systems, computer vision with pre-trained deep learning models (high level), language data, text preprocessing, embeddings, topic modeling, time series, right-censored data / survival analysis
- Part 3: Communication, Ethics, and Deployment

What we didn't cover

- How do these models work under the hood



What would I do differently?

Lots of room for improvement. Here are some things on my mind.

- Balance the pace of the course a bit more (too intense at the beginning, too relaxed at the end!)
- Flipped classroom in a more effective way in the first part of the course.
- Add more interactive components in the lectures
- More activities during lecture time
- Add a course project !?!
- Some material to cover: dealing with outliers, data collection, newer methods

What next?

If you want to further develop your machine learning skills: - Practice! - Work on your own projects - Work hard and be consistent.

- If you are interested in research in machine learning
  - Take CPSC 340. If you do not have the required prereqs you can try to audit it.
- Get into the habit of reading papers and replicating results

? ? Questions for you

For each of the scenarios below

- Identify if ML is a good solution for a problem.
- If yes
  - Frame the problem to a ML problem.
  - Discuss what kind of features you would need to effectively solve the problem
  - What would be a reasonable baseline?
  - Which model would be a suitable model for the given scenario?
  - What would be the appropriate success metrics.

? ? (Practice) Questions for you

App	Goal
QueuePredictor app	Inform callers how long they'll wait on hold given the current call volume
To-doList App	Keep track of the tasks that a user inputs and organize them by date
SegmentSphere App	To segment customers to tailor marketing strategies based on purchasing behaviour
Video app	Recommend useful videos
Dining app	Identify cuisine by a restaurant's menu
Weather app	Calculate precipitation in six hour increments for a geographic region
EvoCarShare app	Calculate number of car rentals in four increments at a particular Evo parking spot
Pharma app	Understand the effect of a new drug on patient survival time

Conclusion & farewell

That's all, folks. We made it! Good luck on your final exam! When you get a chance, please let me know what worked for you and what didn't work for you in this course.