Course Retrospective

Matthew Engelhard

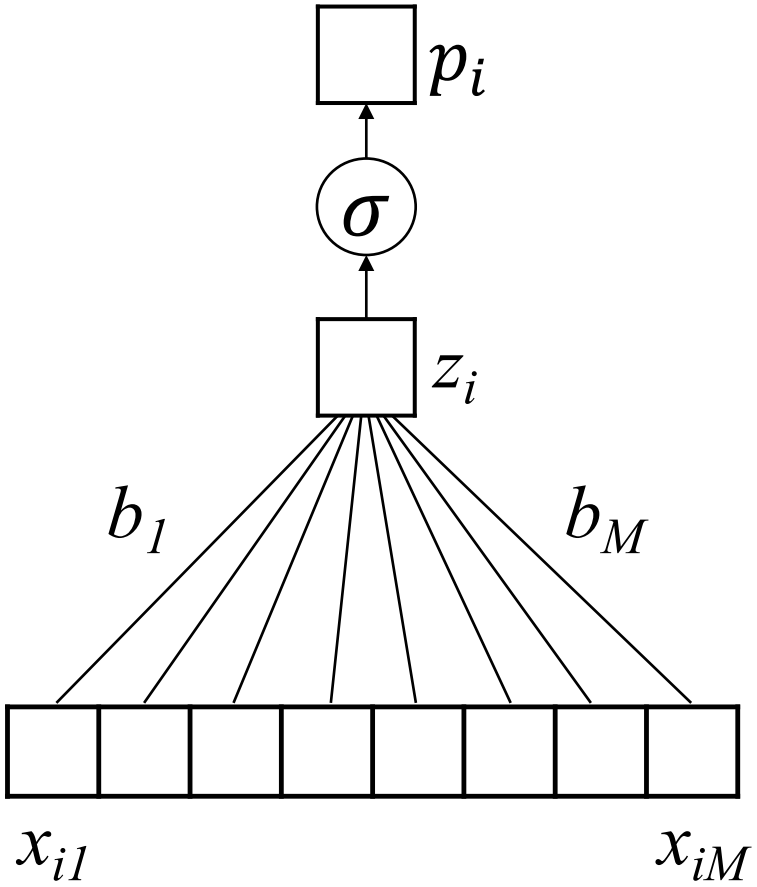
We have covered a lot!

What is machine learning and what problems can it solve?

Logistic regression, MLP, computer vision, NLP

Logistic Regression:

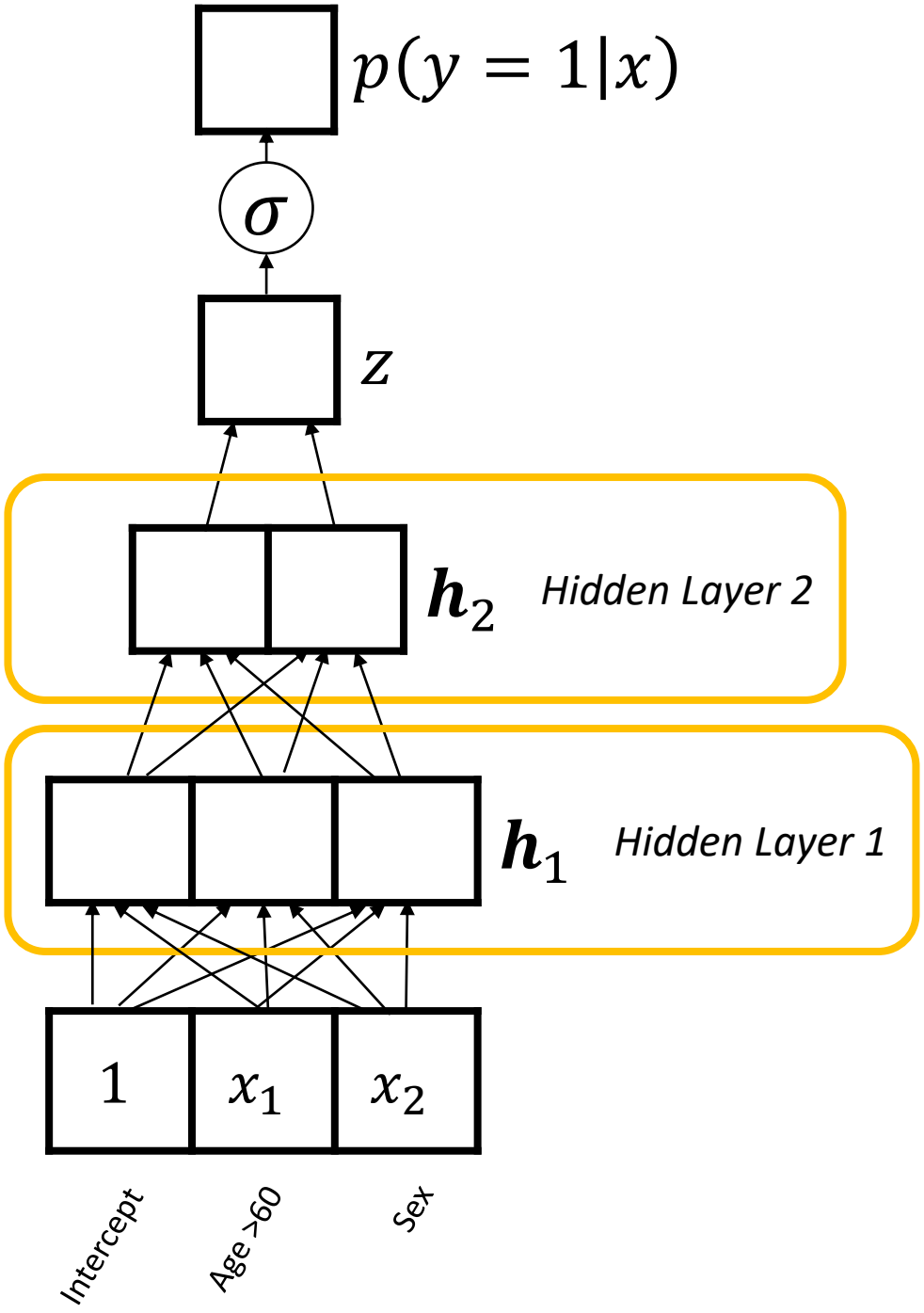
still going strong; use the simplest model that works well



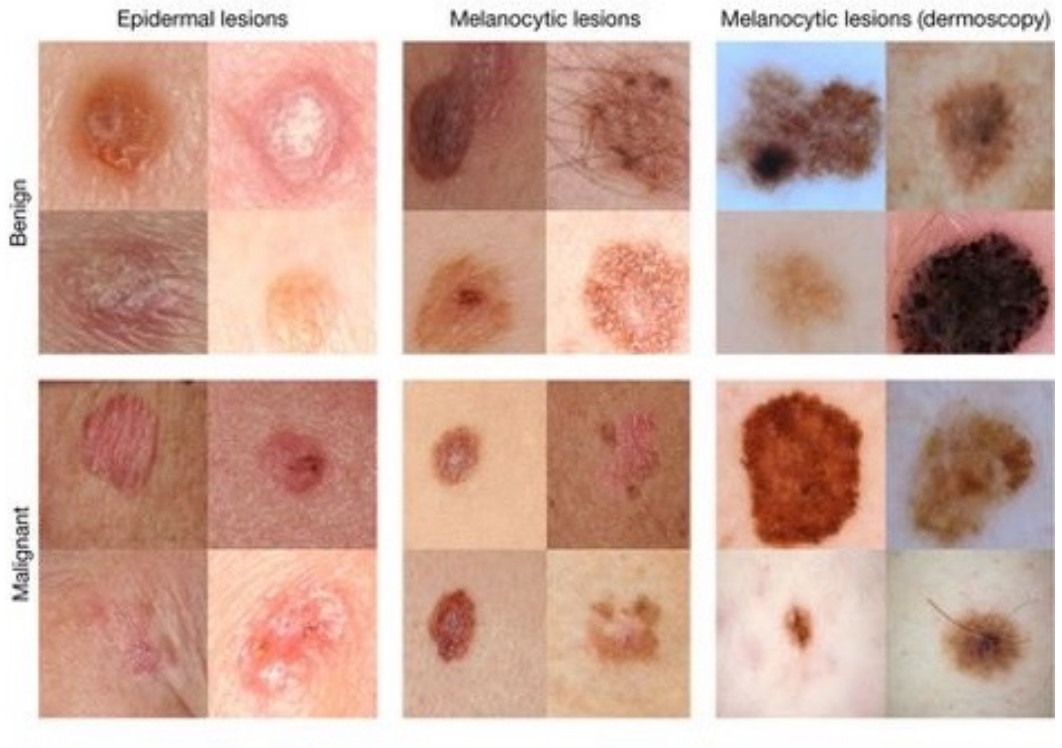
$$p_i = \sigma(b_1 x_{i1} + b_2 x_{i2} + \cdots + b_M x_{iM})$$

MLP for clinical and EHR variables:

what are the advantages?



Overview of medical image processing:
how it works, what it can do

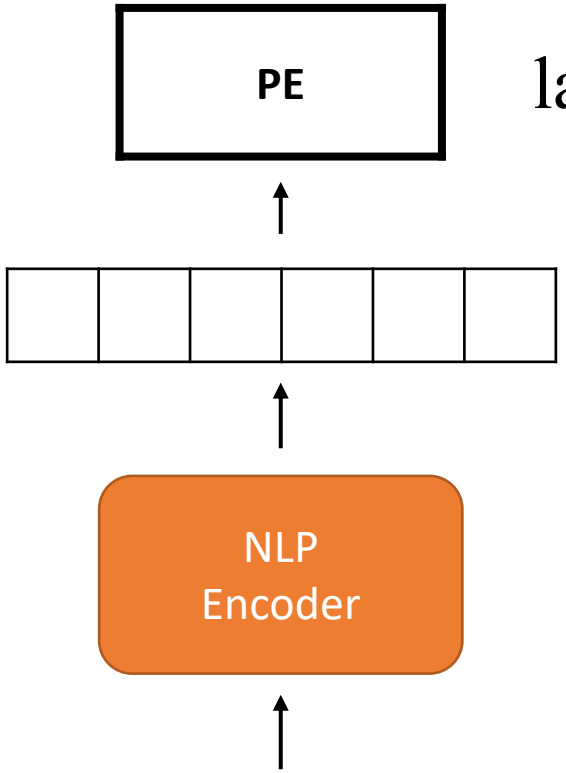


Dermatologist-level classification of skin cancer

Nature volume 542, pages 115–118 (02 February 2017)

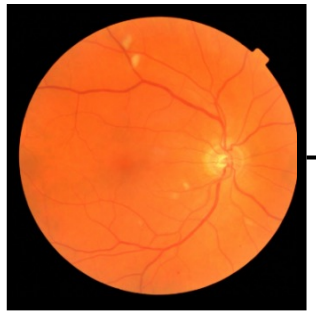
Bag of words & deep NLP

how do we feed text into a predictive model?

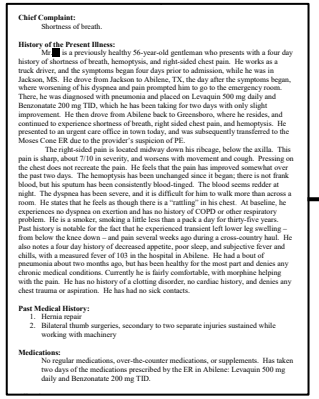


Multi-Modal Prediction Models

1. Train independently
2. Concatenate
3. **Consider adding one more fully-connected, “interaction” layer**
4. Train together
5. **Fine-tune the whole model**



image



Chief Complaint:
Shortness of breath.

History of the Present Illness:
Mr. [redacted] is a previously healthy 56-year-old gentleman who presents with a four day history of shortness of breath, hemoptysis, and right-sided chest pain. He works as a truck driver, and the symptoms began four days prior to admission, while he was in Jackson, MS. He drove from Jackson to Abilene, TX, the day after the symptoms began, where worsening of his dyspnea and pain prompted him to go to the emergency room. There, he was diagnosed with pneumonia and placed on Levaquin 500 mg daily and Benzonatate 200 mg TID, which he has been taking for two days with only slight improvement. He then drove from Abilene back to Greensboro, where he resides, and continued to experience shortness of breath, right sided chest pain, and hemoptysis. He presented to an urgent care office in town today, and was subsequently transferred to the Moses Cone ER due to the provider's suspicion of PE.

The right-sided pain is located midway down his ribcage, below the axilla. This pain is sharp, about 7/10 in severity, and worsens with movement and cough. Pressing on the chest does not recreate the pain. He feels that the pain has improved somewhat over the past two days. The hemoptysis has been unchanged since it began; there is not frank blood, but his sputum has been consistently blood-tinged. The blood seems redder at night. The dyspnea has been severe, and it is difficult for him to walk more than across a room. He states that he feels as though there is a "rattling" in his chest. At baseline, he experiences no dyspnea on exertion and has no history of COPD or other respiratory problem. He is a smoker, smoking a little less than a pack a day for thirty-five years. Past history is notable for the fact that he experienced transient left lower leg swelling – from below the knee down – and pain several weeks ago during a cross-country haul. He also notes a four day history of decreased appetite, poor sleep, and subjective fever and chills, with a measured fever of 103 in the hospital in Abilene. He had a bout of pneumonia about two months ago, but has been healthy for the most part and denies any chronic medical conditions. Currently he is fairly comfortable, with morphine helping with the pain. He has no history of a clotting disorder, no cardiac history, and denies any chest trauma or aspiration. He has had no sick contacts.

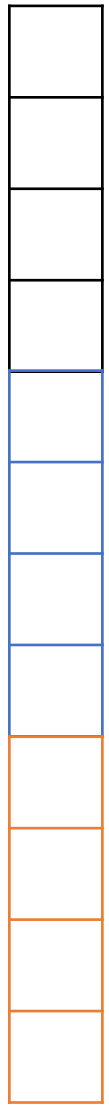
Past Medical History:
1. Hernia repair
2. Bilateral thumb surgeries, secondary to two separate injuries sustained while working with machinery

Medications:
No regular medications, over-the-counter medications, or supplements. Has taken two days of the medications prescribed by the ER in Abilene: Levaquin 500 mg daily and Benzonatate 200 mg TID.

document

CNN Encoder

NLP Encoder



concatenated, multi-modal features

logistic regression

outcome

Concepts needed to participate in model development and critically evaluate ML models

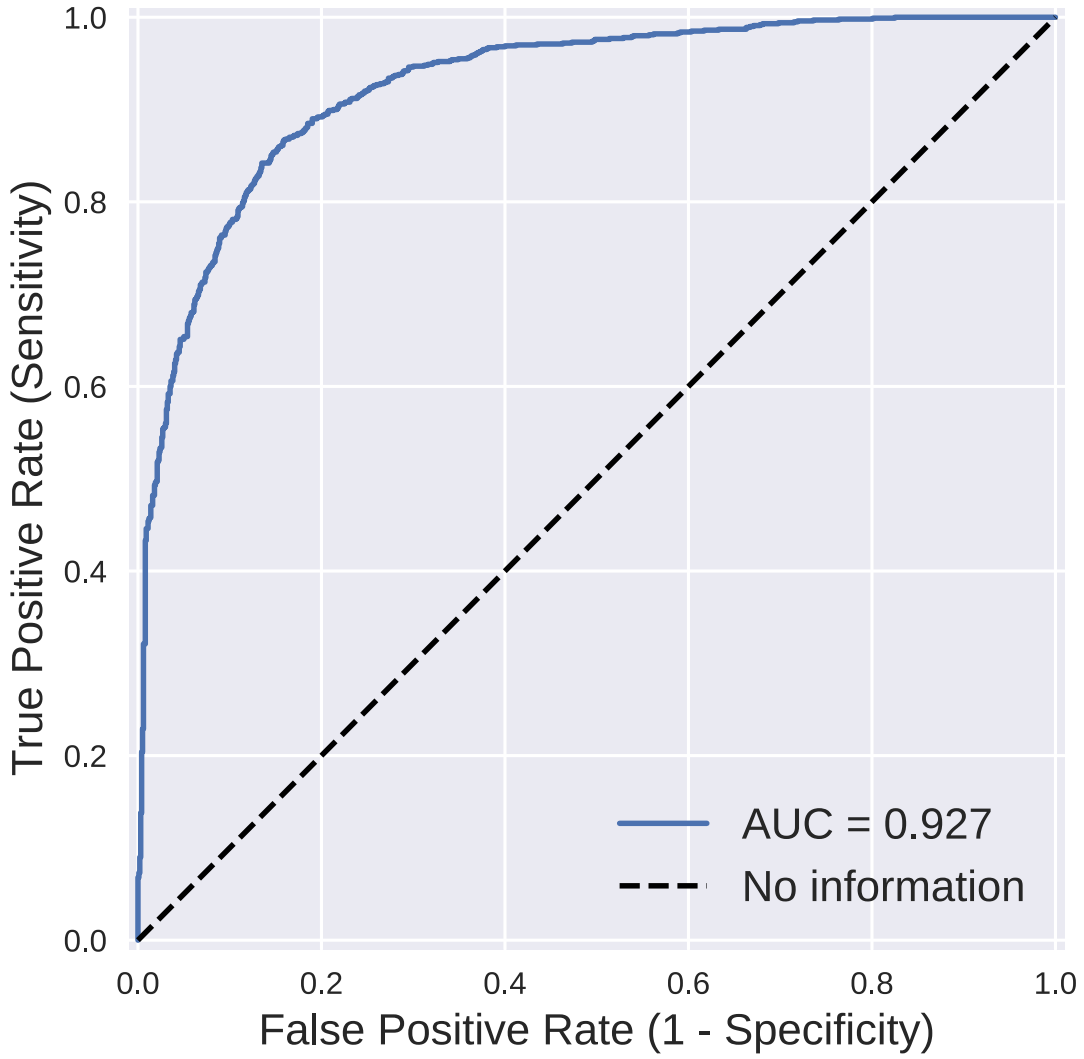
Performance measures; model learning; model development process; mitigating overfitting; understanding model predictions

Performance Measures for Binary Classification

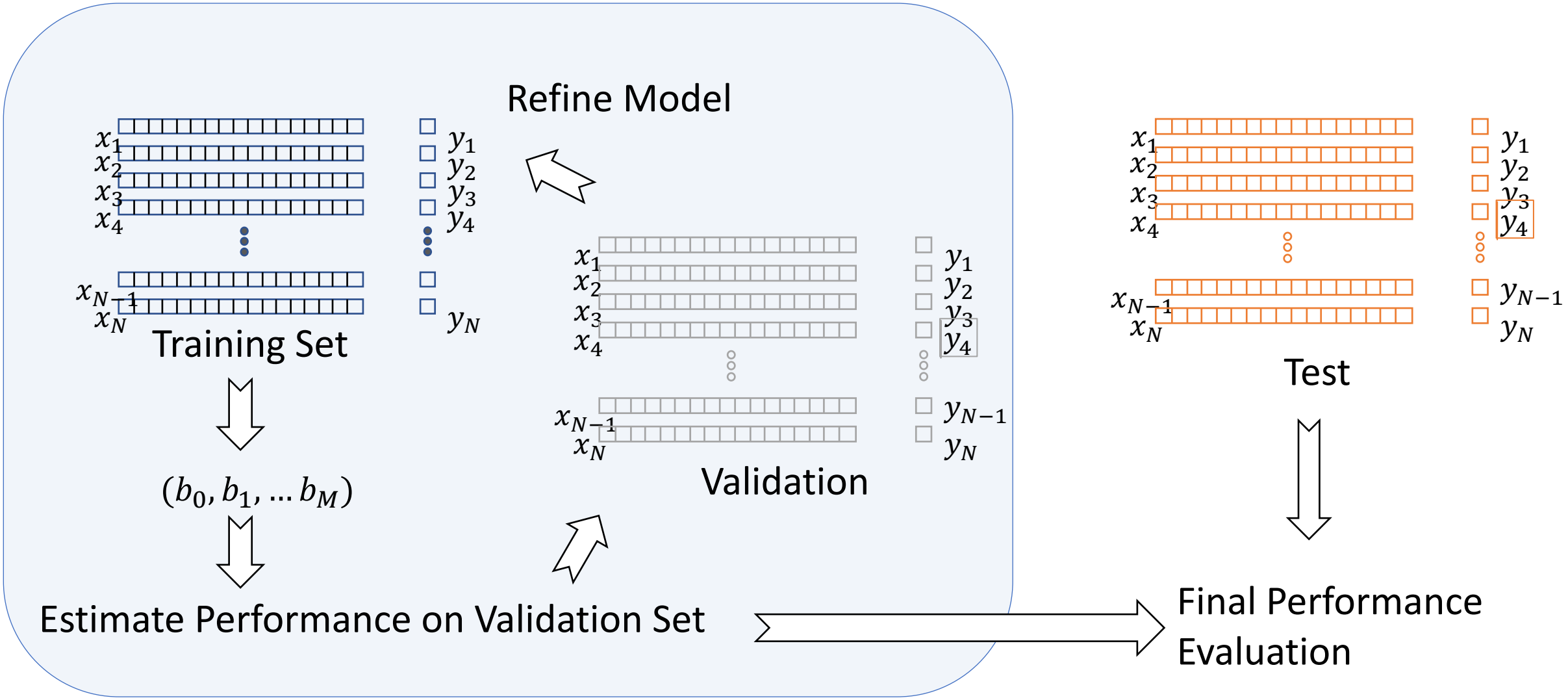
- ROC Curve and AUC
- PR Curve and AP

- Trade-offs between sensitivity, specificity, PPV, NPV

- Choose clinically relevant measures



The Model Development Process



Overfitting and techniques to mitigate it

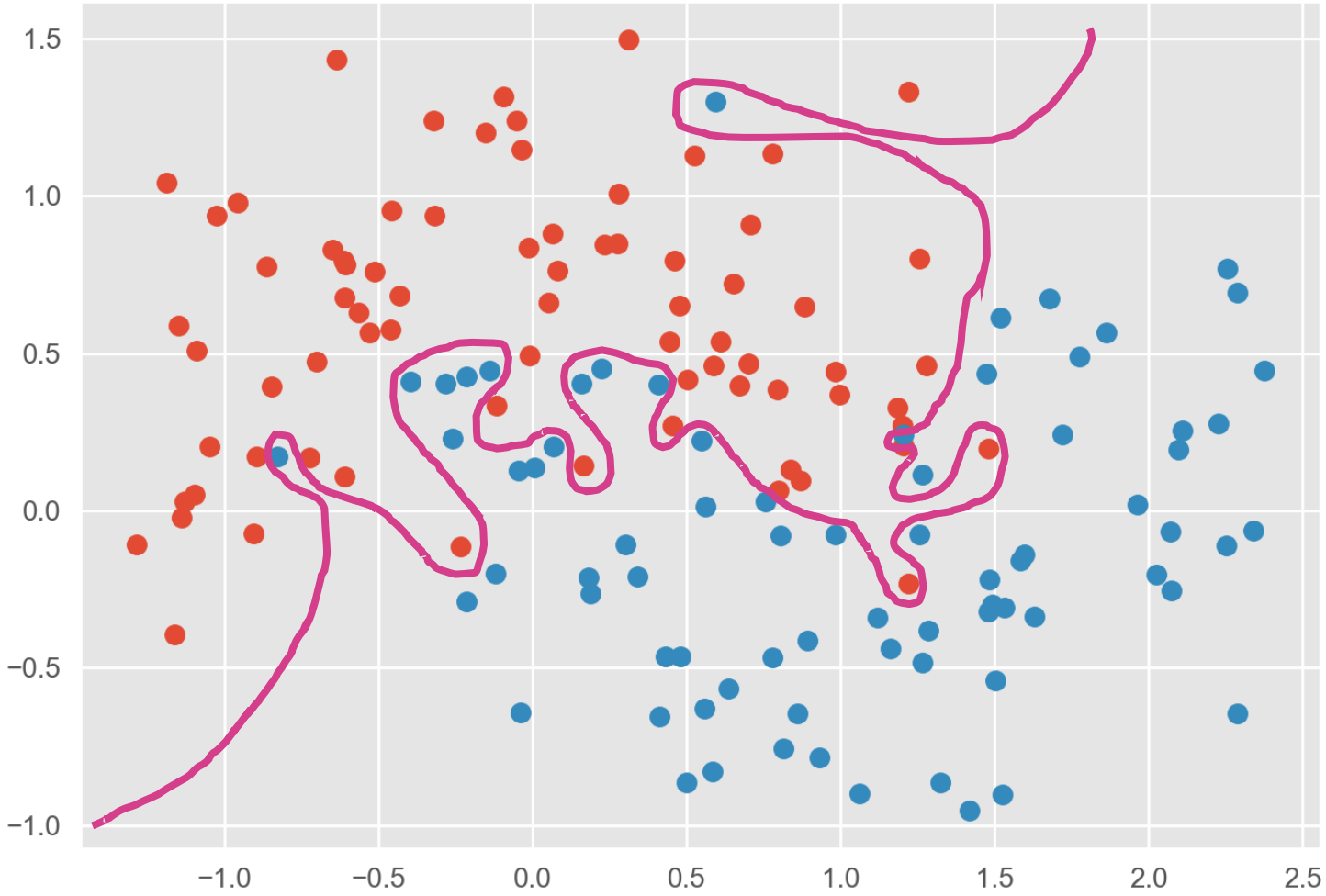
why is this important?

Strategy 1:
Penalize complexity
(regularization)

We can quantify how curvy the line is and add that to the loss.

Strategy 2:
Keep checking the validation set
(early stopping)

When our boundary no longer works well on new data, we know we've gone too far.



Understanding Model Predictions

Course Objective

Understand of the capabilities and limitations of healthcare data science well enough to:

(a) design and manage data science research and/or QA/QI projects

(b) collaborate and communicate effectively with data scientists

(c) critically evaluate data science models and methods, with an emphasis on rigorous model validation

Be in touch: m.engelhard@duke.edu

Thank you!