**Machine Learning Deep Dive**

Delta Analytics builds technical capacity around the world.

This course content is being actively developed by Delta Analytics, a 501(c)3 Bay Area nonprofit that aims to empower communities to leverage their data for good.

Please reach out with any questions or feedback to inquiry@deltanalytics.org.

Find out more about our mission here.

Module 2: Machine learning building blocks.

Course overview:

✓ Module 1: Introduction to Machine Learning ✓ Module 2: Machine Learning Deep Dive ❏ Module 3: Model Selection and Evaluation ❏ Module 4: Linear Regression ❏ Module 5: Decision Trees ❏ Module 6: Ensemble Algorithms ❏ Module 7: Unsupervised Learning Algorithms ❏ Module 8: Natural Language Processing Part 1 ❏ Module 9: Natural Language Processing Part 2

Now let’s turn to the data we will be using...

Module Checklist

❏ Model Development

❏ Defining the machine learning task ❏ Measuring performance of your model ❏ Supervised vs. unsupervised learning methods ❏ Model Validation

Modeling Phase

Now we have our research question, we are able to start modeling!

Data

Exploratory

Research Cleaning

Analysis

Question

Modeling Phase

Task

Learning Research question from Module 1:

Methodology

● How does loan amount requested vary by town?

Performance

Now we have our research question, we are able to start modeling!

Data

Exploratory

Research Cleaning

Analysis

Question

Modeling Phase

Task

We are here!

Performance We will discuss model performance in the next module In this module we introduce the first two steps of modeling:

Learning Methodology

● Defining the machine learning task

● Understanding how the machine learns

Modeling Phase

Let’s start at the basics. Why do we want to build a model?

Source: Deep Learning Book - Chapter 5: Introduction to Machine Learning

Machine learning allows us to tackle tasks that are too difficult to code all possible approaches to on our own.

**By allowing machines to learn from experience**, we avoid the need for humans to specify all the knowledge a computer needs.

**Human Intuition Machine Learning**

**Model**

Based on our experience of the world, we have an understanding of relationships between features

Machine learning models quantify and learn the patterns we observe in data.Computers acquire human intuition and quantify it, by extracting patterns from raw data

All Modeling

Phase

models have 3 key components: a task, a learning methodology and a performance measure

Task What is the problem we want our

model to solve?

Learning Methodology

ML algorithms can be supervised or unsupervised. This determines the learning methodology.

Performance

Quantitative measure we use to Measure

evaluate the model’s performance.

Source: Deep Learning Book - Chapter 5: Introduction to Machine Learning

Today we are looking closer at each component of the framework:

Task What is the problem we want our

model to solve?

What function will map our x (input) Defining f(x)

as close as possible to the true Y (output).

Feature engineering & selection

What assumptions does our model make about the data? Do we have to transform the data? What is x? How do we decide what explanatory features to include in our model?

Is our f(x) correct for this problem?

Learning methodology: how does the model learn the function that best maps x to the true Y?

Learning Methodology

Is our model supervised or unsupervised; how does that affect the learning processing?

How does our ML model learn?

What is our loss function?

Overview of how the model teaches itself.

Every supervised model has a loss function it wants to minimize.

Optimization process

How does the model minimize the loss function.

Performance: How do we evaluate how useful the model is, and how we can improve it?

Performance Quantitative measure we use to

evaluate the model’s performance.

Measures of performance

R2, Adjusted R2, MSE

Feature Performance

Statistical significance, p-values

Ability to

Overfitting, generalize to

underfitting, unseen data

bias, variance

Task and learning methodology

Hold on! Important disclosure:

For the next few slides, we will be introducing the intuition behind machine learning models using **supervised learning** examples. Later in this module, we will explore how unsupervised learning is different.

ML Algorithm

Labelled Data (we have Y in our data)

No Labelled Data

(we don’t have Y in our data)

Supervised learning

Unsupervised

learning

1. Task

Task What is the problem we want our

model to solve?

Research Question Recap: How does loan amount

requested on Kiva vary by town in Kenya?

We have KIVA data about loan amount requested by borrowers all over Kenya.

We want to know how the loan amount requested varies by town.

Building a model involves turning your Task

research question into a machine learning question.

Research question Machine learning

task

How does loan amount requested on Kiva vary by town? ??

Firstly, let’s establish a common vocabulary to talk about the data.

Features

Task

Location.town is an example of a feature. Every column in our data set is a feature!

Source: Subset of Kenyan loans dataset.

Every row of our dataset is an observation. When we include the observation in our model it is part of our training set.

A Task

machine learning task has explanatory features and an outcome feature.

Explanatory features Outcome feature An outcome feature is the **feature we expect to change** when the explanatory features are manipulated. In this example, we expect the loan amount to change when we change the location.

Town borrower lives in

Loan amount requested

What would the outcome features and explanatory features be in the research questions below?

*Try identifying some:*

● What will the price of a stock be tomorrow?

● Does this patient have malaria?

● Would this person buy a car?

The outcome feature might be regression (e.g. $12) or a classification (e.g. Yes or No). We’ll talk about this Solutions:

more later!

**Explanatory features Outcome feature**

Price of a stock market index today Company X’s stock price tomorrow

Age, symptoms, travel history Whether or not a patient has malaria

Income, location Whether or not a person would buy a car

Let’s Task

define our explanatory and outcome features for this task

***Problem*:** I am the mayor of a 30,000 person town and need to justify spending budget on mosquito nets.

I want evidence on how the number of mosquito nets affects the number of cases of malaria. *Can you help?*

Let’s 1.Research

start by identifying the Question

research question!

The research question is what we want to find out from the data, formally stated.

2.Task

How does the number of cases of malaria change when the number of mosquito nets changes?

Next Task

let’s define our task!

1 2 3

Define explanatory and outcome feature

Define f(x) Bring it all

together

explanatory feature(s) outcome feature

**x** Number of

mosquito nets

**Y**Number of people with malaria 2007: 1000 2008: 2200 2009: 6600 2010: 12600

2007: 80 2008: 40 2009: 42 2010: 35

We also call our explanatory features x, and our outcome feature Y. **Looks like as mosquito nets increase, the number of malaria cases decreases.**

Define explanatory and outcome feature

How does the number of cases of malaria change when the number of mosquito nets changes?

What Task

would you conclude from looking at this data? How many nets would you recommend? x Number mosquito of

nets

Y

Number of people with malaria 2007: 1000 2008: 2200 2009: 6600 2010: 12600

2007: 80 2008: 40 2009: 42 2010: 35

You came to a conclusion by **recognizing a pattern in the data**. This is similar to how a machine learning algorithm would approach the same problem.

Machine Task

learning allows us to learn from historical patterns.

If Mr. Mayor had **no** machine learning methods to use, he could find an answer by trying a different # of nets year after year.

But this has an obvious **human cost**, and it would be very hard to update the model to account for, for example, new residents to his town.

Machine learning algorithms help answer questions without this human cost - we are **learning from data**, or in other words, **learning from history**!

**Human Intuition Machine Learning Model**

“Over four years, increasing number of mosquito nets decrease the number of malaria cases.”

An increase in x (mosquito nets) **causes** a decrease in Y (malaria cases).

● Humans form rules based upon observation and pattern recognition.

● ML model takes input x and maps it to the output Y.

Our model f(x) is a function that maps our input x to a predicted Y\*.

irreducible error

Define f(x)

explanatory feature(s) model predicted outcome

x Number of

mosquito nets Predicted Number of

people with malaria Y\* f(x) + e

The Define f(x)

goal of f(x) is to predict a Y\* as close to the true Y as possible.

My job is to make

Our function f(x) maps an input x to the predictions as

a predicted Y, which we refer to as **useful** as possible!

Y\*. We want to choose an f(x) that will map x as close to the true Y as possible.

e is irreducible error. This captures error caused by factors like measurement error, randomness in the data, and inappropriate model choice. No matter how well you optimize your model, this will never be reduced to 0.

Predicted Number of people with malaria f(x) + e = Y\*

We want Y\* to be close to true Y because we want the function to output useful predictions.

Define f(x)

Y\*=f(x)+e

In this example, predicted Y appears close to the true Y. We will talk about how to quantify this in the next section.

True Y

We want Y\* to be close to true Y because we want the function to output useful predictions.

Define f(x)

Y\*=f(x)+e

In this example, predicted Y appears far from the true. ***This is probably not very useful***. We will talk about how to quantify this in the next section.

True Y

What is f(x)? It depends on the machine Define f(x)

learning algorithm we choose. **x f(x) Y\***

explanatory feature(s), like number of mosquito nets

predicted outcome, e.g. Number of people with malaria

**Supervised learning algorithms:**

- Linear regression - Decision tree - Random forest - ...

Examples of f(x):

**Unsupervised learning algorithms:**

- K-means clustering - Hierarchical clustering - ...

When you have labelled data When you don’t have labelled data

Research question

Machine learning task

x Number of

mosquito nets

Bring it all together

Let’s bring everything together.

How does the number of cases of malaria change when the number of mosquito nets changes?

f(x)

Number of people with malaria Y\*

The task function depends upon the data type you want to predict. Supervised learning problems fall into two main categories: regression & classification.

Supervised learning task

A regression problem is when we are trying to **predict a numerical value**, such as “dollars” or “weight”.

Source: Andrew Ng, Stanford CS229 Machine Learning Course

**Regression**

Continuous variable

A classification problem is when we are trying to **predict whether something belongs to a category**, such as “red” or “blue” or “disease” and “no disease”.

**The Task**

**Classification**

**Classification**

Categorical variable

2. Learning Methodology

Learning Methodology

ML algorithms can be supervised or unsupervised. This determines the learning methodology.

Learning methodology: how does the model learn the function that best maps x to the true Y?

Learning Methodology

Is our model supervised or unsupervised? How does that affect the learning processing?

How does our ML model learn?

What is our loss function?

Overview of how the model teaches itself.

Every supervised model has a loss function it wants to minimize.

Optimization process

How does the model minimize the loss function?

Recall that machine learning is a subset of data science that allows machines to learn from raw data.

Machine learning

Traditional software programing involves giving machines instructions which they perform. **Machine learning involves allowing machines to learn from raw data so that the computer program can change when exposed to new data (learning from experience).**

Source: https://www.youtube.com/watch?v=IpGxLWOIZy4

Learning Methodology What do we mean when we say a machine

“learns from experience”?

Machine learning is a subset of data science that allows machines to learn from raw data.

How does the model learn from raw data?

How Learning

the algorithm learns depends upon Methodology

type of data you have.

START: Research question

Do I have data that may answer my question?

Gather or find more data Do you have labelled data?

Unsupervised learning methods

You are here!

N

Y N

Y Supervised learning methods

What Learning

does labelled data mean? Methodology

Yes

Y=Number of people

The outcome feature (Y) you are interested in

with malaria

2007: 80 predicting is recorded in

2008: 40

Do you have labelled data?

the data. If you have a labelled Y, you can use supervised learning

2009: 42 2010: 35

methods.

No

Y=Number of people with malaria

The outcome feature (Y) is not recorded in the

2007: 2008: data. You do not have a

2009: labelled Y.

2010:

Learning Methodology

Do you have labelled data?

Whether or not you have labelled data determines whether it is a Supervised or Unsupervised Learning problem

f(x) + e = Y\*

Y is in your data

Unsupervised Learning

*●* For every x, there is no Y

Y is not in your data

Supervised Learning Yes

● For every x, there is a Y

● Goal is to **predict** Y using x

No

● Goal is not to predict, but to **investigate** x

Most problems you will initially encounter are supervised algorithms. *How do supervised algorithms learn?*

Intuitive supervised learning

algorithms explanation learn: for how supervised

Y

Number with malaria of people

2007: 80 2008: 40 2009: 42 2010: 35

Imagine you are a teacher and you ask your students a question.

The labels Y provide the correct answer for the problem the students are trying to solve. **Since you know the correct answer, you can reward good student performance and punish poor performance**. This encourages ongoing learning!

Extending supervised learning

researcher) this are example, the teacher you (the

and the Model is the student.

Model

I want to get the correct

Great Mr. Model! answer for predicting Y

Once you give me and be the best student in

your answer I will the class.

let you know the correct answer.

Every time Mr. Model predicts Y\*, you compare Y\* to the true Y to see how well he did.

Our supervised learning

Model starts trying to provide an estimated Y\* by guessing. Model I have never seen this

problem before! I’ll just start by randomly guessing an answer

Y\* Y Predicted Number of

Actual Number of people with malaria

people with malaria and see what happens.

2007: 1

2007: 80 2008: 2000

2008: 40 2009: 300

2009: 42 2010: 40

2010: 35

Unsurprisingly, the results appear terrible, judging from the fact that actual numbers are very different from the predicted numbers. ***To quantify how bad or good results are, we use Y-Y\*.***

Which Learning

model is more useful at Methodology

mapping x close to true Y?

Y\*=f(x)+e

What is our loss function?

a) b)

True Y

Which prediction was worse, a) or b)?

True Y

Y\*=f(x)+e

True Y

Y\*=f(x)+e

What is our loss function?

Learning Methodology

We can immediately tell that b is better!

We can immediately tell that b is better!

Y\*=f(x)+e

We can see that the f(x) in b maps x to a Y\* much closer to the true Y. A **loss function** allows us to quantify this difference.

a) b)

True Y

A model’s goal is to minimize the loss function.

We have already seen one simple example: Y-Y\*, or the difference between the predicted Y and the actual Y. Later, we will see more sophisticated loss functions.

Source: Stanford ML Lecture 1

**The Task**

**The Loss Function**

A loss function quantifies how unhappy you would be if you used f(x) to predict Y\* when the correct output is y. It is what we want to minimize.

Another way to think about it is that a loss function quantifies how well our f(x) fits our data.

Another way to think about it is that a loss function quantifies how well our f(x) fits our data.

supervised learningWe first decide on how to measure how unhappy we are with these results. We call this our **loss function**. On the next slide, we show a few different possible loss functions we can use to assess Mr. Model.

Y\* Y Since I know the right answer, I can compare predicted Y\* to the

Predicted people with Number malaria

of

Actual people with Number malaria

of

true Y to help guide

2007: 1

2007: 80 Mr. Model.

2008: 2000

2008: 40 2009: 300

2009: 42 2010: 40

2010: 35

Recall that there are two different types of tasks: Supervised learning task

A regression problem is when we are trying to predict a numerical value given some input, such as “dollars” or “weight”.

Source: Andrew Ng, Stanford CS229 Machine Learning Course

**Regression**

Continuous variable

A classification problem is when are trying to predict whether something belongs to a category, such as “red” or “blue” or “disease” and “no disease”.

**The Task**

**Classification**

**Classification**

Categorical variable

The Learning Methodology

What is our loss function?

choice of loss function depends upon the type of task. We will discuss loss functions for both types of task.

**Regression**

absolute error (L1)

**Classification**

log loss

least squares error

Continuous variable

(L2)

Categorical variable

hinge loss

Source: Andrew Ng, Stanford CS229 Machine Learning Course

Including mean squared error (MSE), root mean squared error (RMSE)

There Learning Methodology

What is our loss function?

are a few different loss functions we could choose from, depending on the problem we are trying to solve. **Regression Classification**

Continuous variable Categorical variable

absolute error

root mean squared (L1)

error (RMSE)

log loss

least squares error

mean squared error (L2)

(MSE)

hinge loss

**Regression loss functions**

1. L1 norm (mean absolute error) 2. L2 norm (least squares error)

- Mean squared error

What Learning

is our Methodology

loss function?

Outcome feature in data is continuous

Y

Number of people with malaria 2007: 80 2008: 40 2009: 42 2010: 35

Regression task

Our outcome feature is continuous: the number of people who have malaria.

L1 or L2 loss function

Takes the mean of the L2 loss over all observations.

Source: L1 and L2

Learning Methodology

L1 What is our

and L2 loss are two possible options loss function?

for assessing how unhappy we are with Mr. Model’s choice of f(x).

absolute error (L1)

least squares (L2) error

mean squared error

Also called L1 loss, this minimizes the **sum** of absolute errors between True Y and predicted Y\*.

Also called L2 loss, this minimizes the **square** of the error between True Y and predicted Y\*.

How Learning Methodology

bad were Mr. Model’s initial results? Let’s compute the L1 norm. How well did Mr. Model's random guess perform?

Source: Intro to Stat - Introduction to Linear Regression

Y\* Predicted Number of

Y Actual Number of people with malaria

people with malaria

2007: 1 2008: 2000 2009: 300 2010: 40

2007: 80 2008: 40 2009: 42 2010: 35

absolute error (L1)

What is our loss function?

(|1-80|+|2000-40|+|300-42|+|40-35|) = 2,302

How Learning Methodology

bad where Mr. Model’s initial results? Let’s compute the L2 norm. How did Mr. Model’s initial random guess do?

Source: Intro to Stat -Introduction to Linear Regression

Y\* Y Predicted people with Number malaria

of

Actual people with Number malaria

of

2007: 1

2007: 80 2008: 2000

2008: 40 2009: 300

2009: 42 2010: 40

2010: 35

Least squares error

(L2) (80-1)^2+(40-2000)^2+(42-300)^2

+(35-40)^2=3,914,430

What is our loss function?

Takes the mean of the L2 loss over all observations.

Source: L1 and L2,

Learning Methodology

What is our loss function?

least squares error

(L2) mean squared error

Also called L2 loss, minimizes the square of the error between True Y and predicted Y\*.

We can normalize our L2 loss by computing mean squared error or root mean squared error.

RMSE = sqrt(mean(S)) MSE = mean(S)

root mean squared error

Takes the square root of the mean of the L2 loss.

Mean Learning Methodology

What is our loss function?

squared error takes the average L2 error per observation.

How did Mr. Model’s initial random guess do?

Y\* Y Predicted people with Number malaria

of

Actual people with Number malaria

of

2007: 1

2007: 80 2008: 2000

2008: 40 MSE = mean(S)

2009: 300

2009: 42 2010: 40

2010: 35 ((80-1)^2+(40-2000)^2+(42-300)^2 Mean squared error

+(35-40)^2)/4 = 978,607.5

Source: Intro to Stat -Introduction to Linear Regression

Root Learning Methodology

What is our loss function?

mean squared error takes the square root of the average L2 error per observation.

Source: Intro to Stat -Introduction to Linear Regression

2007: 1 2008: 2000 2009: 300 2010: 40

Root mean squared error How did the models initial random guess do?

Y\* Y Predicted people with Number malaria

of

Actual people with Number malaria

of

2007: 80

RMSE = sqrt(mean(S))

2008: 40 2009: 42 2010: 35 (((80-1)^2+(40-2000)^2+(42- 300)^2+(35-40)^2)/4)^(1⁄2) = 989.25

We Learning Methodology

can compute for each loss functions how unhappy we are with models initial random guess.

Don’t worry about these numbers. What’s important is you understand how we are transforming them step by step.

absolute error (L1) least squares (L2) error

mean squared error

Also called L1 loss, minimizes the sum of absolute errors between True Y and predicted Y\*.

2,302 3,914,430 978,608

989

root mean squared error

Also called L2 loss, minimizes the square of the error between True Y and predicted Y\*.

Takes the average L2 loss per observation in the data.

MSE = mean(S)

RMSE = sqrt(mean(S)) Takes the square root of the average L2 loss per observation in the data.

Source: L1 and L2

RMSE is the square root of the average L2 loss per observation.

MSE

Learning Methodology

This There are five steps to RMSE:

Y-Y\* For every observation in our

dataset, measure the difference job isn’t done

between true Y and predicted Y.

until I reduce RMSE.

^2 Square each Y-Y\* to get the

absolute distance, so positive values don’t cancel out negative ones when we sum.

Sum Sum across all observations so we

get the total error.

Model

mean Divide the sum by the number of

observations we have.

root Take the square root of the mean

calculated above.

Source: Intro to Stat -Introduction to Linear Regression

**Which loss function should we use?**

1. L1 norm (mean absolute error) 2. L2 norm (least squares error)

Each Learning Methodology

What is our loss function?

loss function has important pros and cons.

absolute error (L1) least squares error

(L2)

Source: L1 and L2

vs.

MSE and RMSE are both normalized versions of L2 error. If we decide to use L2, we will choose MSE or RMSE.

**Robust? Stable Solution? How many**

**solutions?**

L1 Robust Not stable Multiple possible

solutions

L2 Not very robust Stable One possible

solution

Learning Methodology

If What is our

loss function?we decide to use least squares

error (L2), we may decide to report RMSE OR MSE

MSE vs. RMSE The key difference between RMSE and MSE is that taking the root in RMSE normalizes the error to the same units of measurement.

This makes the error term more interpretable.

Both MSE and RMSE amplify and severely penalize large errors more than small ones by squaring the error.

Classification loss functions

1. Log loss 2. Hinge loss

Let’s Define explanatory

define a slightly different task so and outcome feature

we can discuss hinge and log loss.

**Task: We want to predict whether or not a patient has malaria using their temperature.**

X Temperature Of patient

Y 39.5°C 37.8°C 37.2°C 37.2°C

Does the patient have malaria?

NoYes Yes No

Learning Methodology

Our outcome feature is categorical: we want to predict whether or not someone has malaria. This is a binary classification problem.

What is our loss function?

What is our loss function?

This is a classification task, so we can use either log loss or hinge loss.

But first, what is a classification task?

Classification Classify based

tasks output the probability of upon probability threshold

belonging to a class. Normally, based upon a threshold of 50% we then assign the predicted class.

outcome feature predicted probability predicted outcome

Y Does the

patient have

Y\* Does the model

predict that the malaria?

patient has malaria?

NoYes Yes

Yes Yes

Yes No

No What is the probability that the patient has malaria?

0.55 0.80 0.85 0.2

We Classify based

can evaluate accuracy by looking just at the upon probability threshold

predicted outcome vs. the actual outcome. Here, accuracy is 75%!

outcome feature predicted probability predicted outcome

Y Does the

patient have

Y\* Does the model

predict that the malaria?

patient has malaria?

NoYes Yes

Yes Yes

Yes No

No What is the probability that the patient has malaria?

0.55 0.80 0.85 0.2

However, **we are missing out on using probability**, which is important information about how certain the model is about its prediction. Let’s look at some loss functions that utilize this metric.

For every prediction the model Makes, we can Log loss

measure the logarithmic loss. What is logarithmic loss?

I’m only 60% sure that the answer is Yes...

Model

Source: https://www.r-bloggers.com/making-sense-of-logarithmic-loss/

Log loss evaluates the probability of belonging in a class

- The smaller the log loss, the smaller the

uncertainty, the better the model

- A perfect classifier would have log loss = 0 - Log loss heavily penalizes classifiers that are

confident about an incorrect classification

- Ways to improve log loss:

- Are there problematic errors in dataset? - Do we want to smooth the probabilities?

For every prediction our model Makes, we can also measure the hinge loss. What is hinge loss? Hinge loss

Hinge loss is the logical extension of the regression loss function, **absolute loss**.

**Absolute loss**: Y- Y\*, where Y and Y\* are integers.

**Hinge loss**: max(0,1-(Y\*)(Y))

Where Y can equal -1 (no) or 1 (yes) for each class.

For each observation, if Y\* == Y (both are 1 or both are -1), hinge loss = 0. If Y =/= Y\*, hinge loss **increases**.

*The cumulated hinge loss is therefore the upper bound of the number of mistakes made by the classifier.*

Sources: https://en.wikipedia.org/wiki/Hinge\_loss; http://scikit-learn.org/stable/modules/generated/sklearn.metrics.hinge\_loss.html

How Learning

What is our

do we choose a loss function for a Methodology

loss function?

classification problem?

Log Loss vs. Hinge Loss Leads to **more exact probabilities**, but at the cost of accuracy

Leads to **better accuracy**, but at the cost of exact probabilities

Learning Methodology

How do we choose a loss function for a classification problem?

What is our loss function?

What is our loss function?

***Depends on the question you want to answer!***

E.g. For a problem where we are trying to assess patient health, we know that *false positives* (the model predicts you do have malaria, but you actually don’t) are safer and generally more preferable than *false negatives* (the model predicts you don’t have malaria, but you actually do.)

Therefore it is probably safer to evaluate our output as a **probability** of whether or not you have malaria. We will use **log loss**.

A Learning note on categorical loss functions ... Methodology

What is our loss function?

*We provide only a broad conceptual overview of log loss and hinge loss as we will not be using them in our coding lab. However, we encourage you to explore them further.*

***More resources can be found at the end of this module.***

Our model computed an initial guess using RMSE. **How does he improve on his initial guess?**

Learning Methodology

Oh no! That wasn’t very good, let’s try something else! 1st initial

guess

If the new f(x) reduces the loss, **our model keeps changing the f(x) in that direction.** After every change, the model measures whether the loss has increased, decreased or stayed the same.

Source: Intro to Stat -Introduction to Linear Regression

What is our loss function?

Our initial RMSE is very high. Our model tries a different f(x) and compares RMSE.

2nd update 3rd update

RMSE 1,000 1,300 800

# nets 300 100 400

As the model updates the prediction at each step, we see that themodel is learning - it changes in response to a higher MSE.