Machine Learning from Data

Lecture 9: Spring 2021

Today’s Lecture

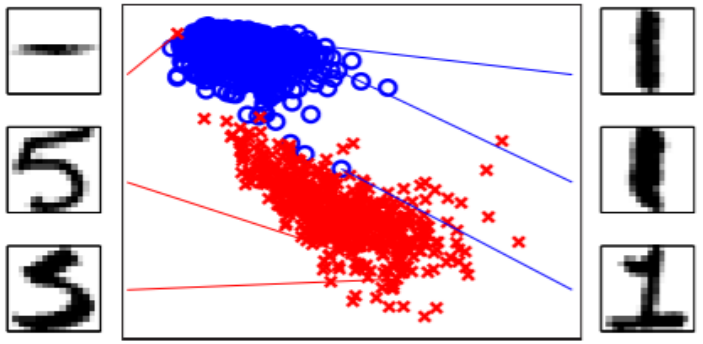
- Logistic Regression
- Gradient Descent

Previous Lecture

The linear signal:

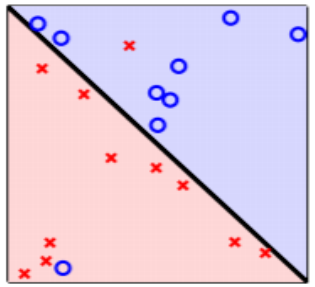
$$s = \mathbf{w}^{\mathrm{T}}\mathbf{x}$$

Good Features are Important

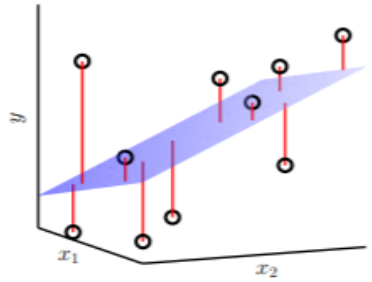


Before looking at the data, we can **reason** that symmetry and intensity should be good features **based on our knowledge of the problem.**

Algorithms



Linear Classification.
Pocket algorithm can tolerate errors
Simple and efficient



Linear Regression.
Single step learning:

$$\mathbf{w} = \mathrm{X}^{\dagger}\mathbf{y} = (\mathrm{X}^{\mathrm{T}}\mathrm{X})^{-1}\mathrm{X}^{\mathrm{T}}\mathbf{y}$$

Very efficient $O(Nd^2)$ *exact* algorithm.

Predicting a Probability

Will someone have a heart attack over the next year?

age	62 years
gender	male
blood sugar	120 mg/dL40,000
HDL	50
LDL	120
Mass	190 lbs
Height	$5' 10''$
...	...

Classification: Yes/No

Logistic Regression: Likelihood of heart attack

logistic regression $\equiv y \in [0, 1]$

Properties of the Sigmoid

•

How to get the Probability?

-

Target function

•

How to minimize E_{in}?

- $$E_{\text{in}}(\mathbf{w}) = \frac{1}{N}\sum_{n=1}^{N} \ln(1 + e^{-y_n \cdot \mathbf{w}^{\mathrm{T}}\mathbf{x}_n})$$

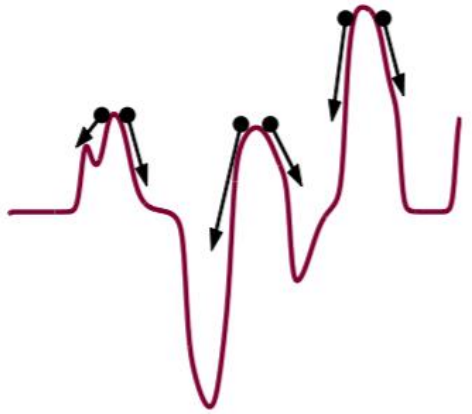
Ball Phenomena

Ball on a complicated hilly terrain

— rolls down to a *local valley*

↑

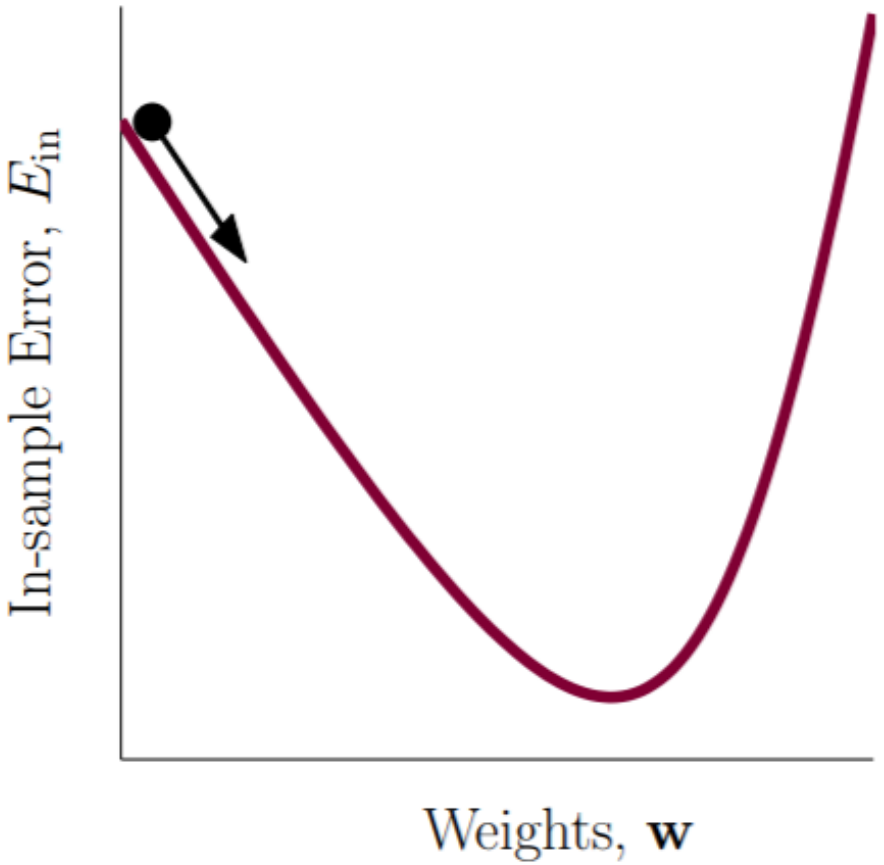
this is called a *local minimum*



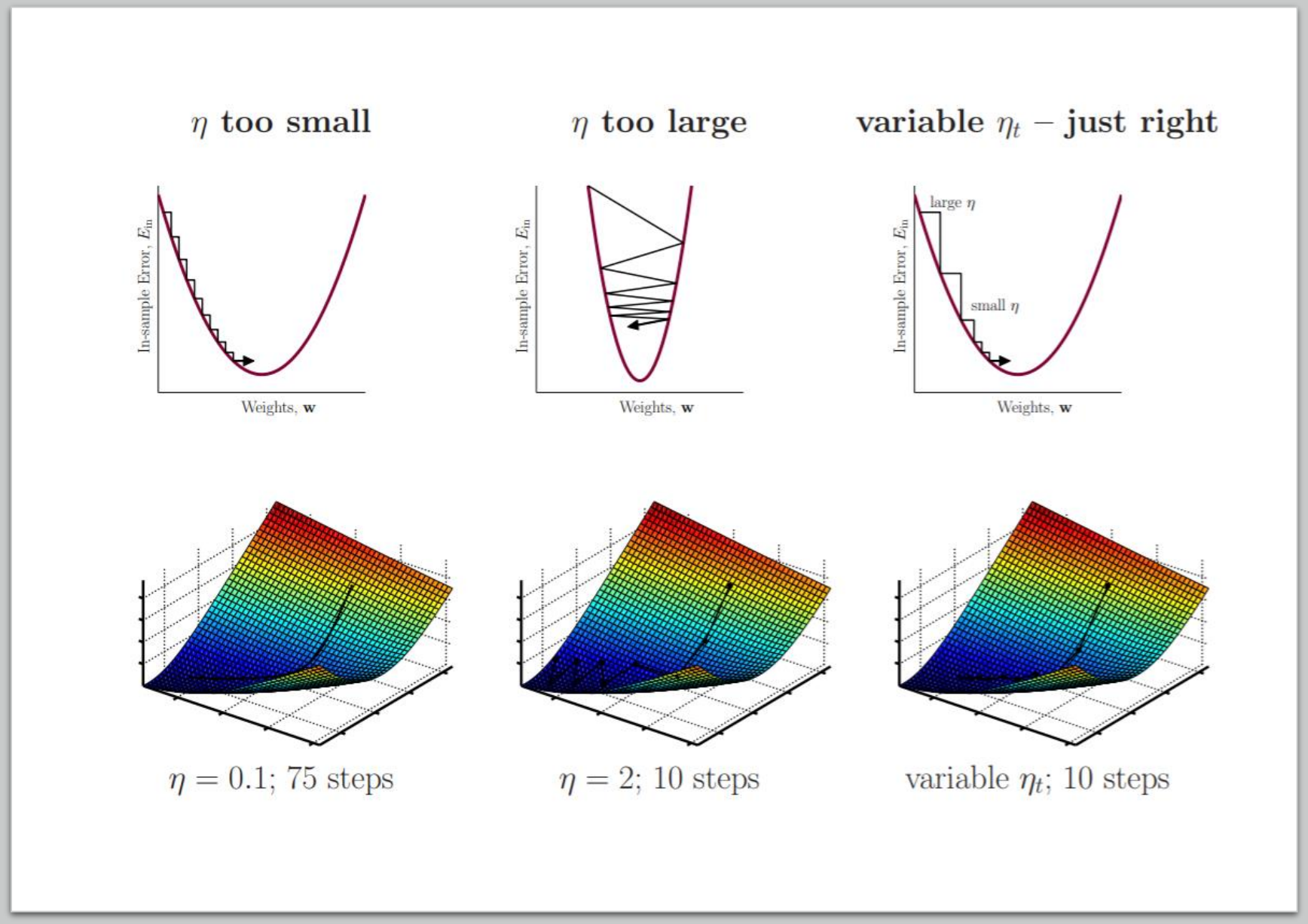
Questions:

How to get to the bottom of the deepest valey?

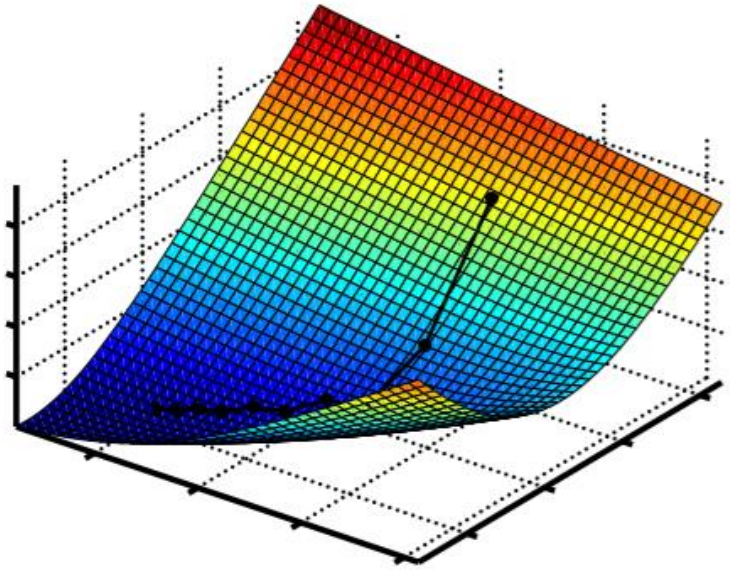
How to do this when we don't have gravity?


In-sample Error, E_{in}
Weights, $\mathbf{w}$

Step Size

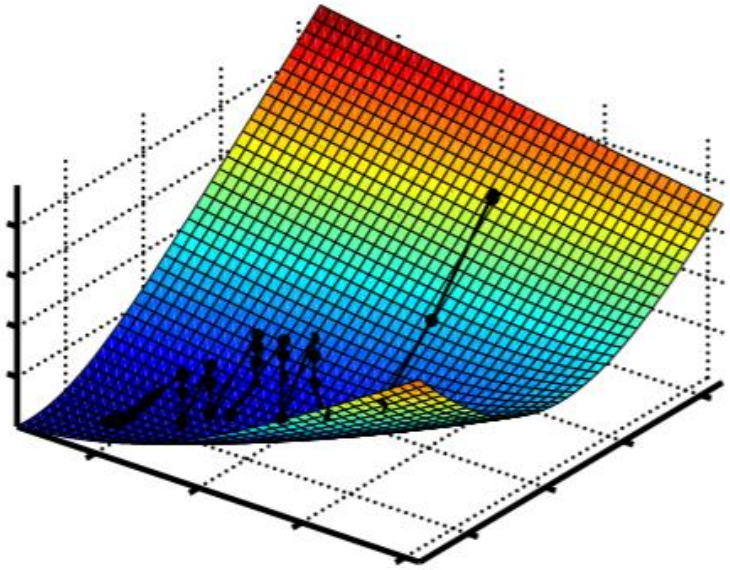

η too small
η too large
variable η_t – just right
In-sample Error, E_{in}
Weights, $\mathbf{w}$
large η
small η
$\eta = 0.1$; 75 steps
$\eta = 2$; 10 steps
variable η_t; 10 steps

GD



$\eta = 6$
10 steps
$N = 10$

SGD



$\eta = 2$
30 steps

Thanks!