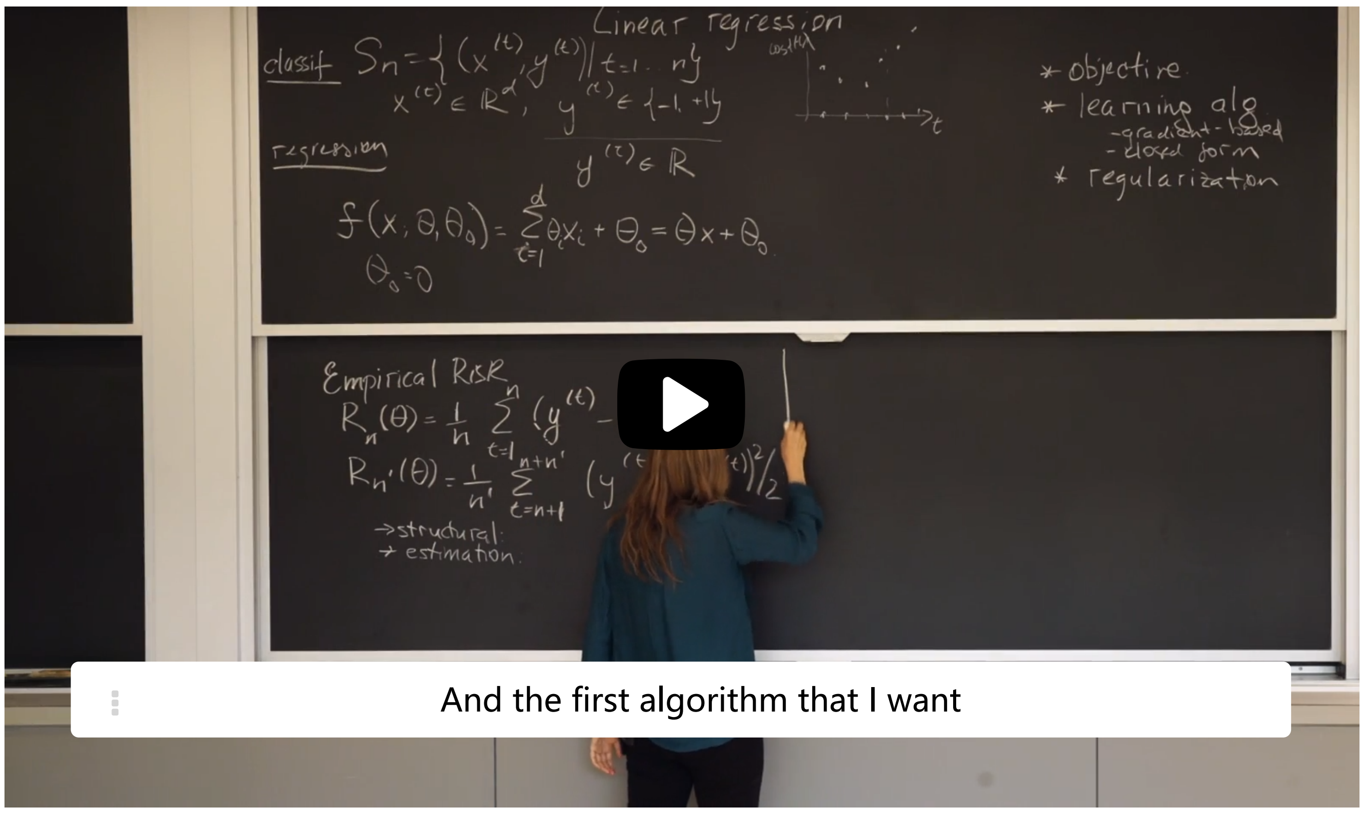
# 5. Gradient Based Approach

## Learning Algorithm: Gradient Based Approach




0:02 / 7:03 1.0x

And the first algorithm that I want

to demonstrate-- so we'll just close where we are,

our objective.

We are done, and now we are talking

about finding algorithm.

And as you can see here, we will start first

with the gradient-based approach.

So the good news about this function,

that it's actually differentiable everywhere,

correct?

Video

Download video file

Transcripts

Download SubRip (.srt) file

Download Text (.txt) file

xuetangX.com 学堂在线

## True or False

1/1 point (graded)

Let $R_n(\theta)$ be the least squares criterion defined by

$$R_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \text{Loss}\left(y^{(t)} - \theta \cdot x^{(t)}\right).$$

Which of the following is true? Choose all those apply.

- [x] The least squares criterion $R_n(\theta)$ is a sum of functions, one per data point. ✔
- [ ] Stochastic gradient descent is slower than gradient descent.
- [x] $\nabla_\theta R_n(\theta)$ is a sum of functions, one per data point. ✔

✔

**Solution:**

For every point, the loss is a function of $\theta$, so the least squares criterion $R_n(\theta)$ is a sum of functions, one per data point, and this is what makes stochastic gradient descent possible. We want to do stochastic gradient descent because it is faster than gradient descent. Finally, because $R_n(\theta)$ is sum of functions, one per data point, $\nabla_\theta R_n(\theta)$ is also a sum of functions one per data point.