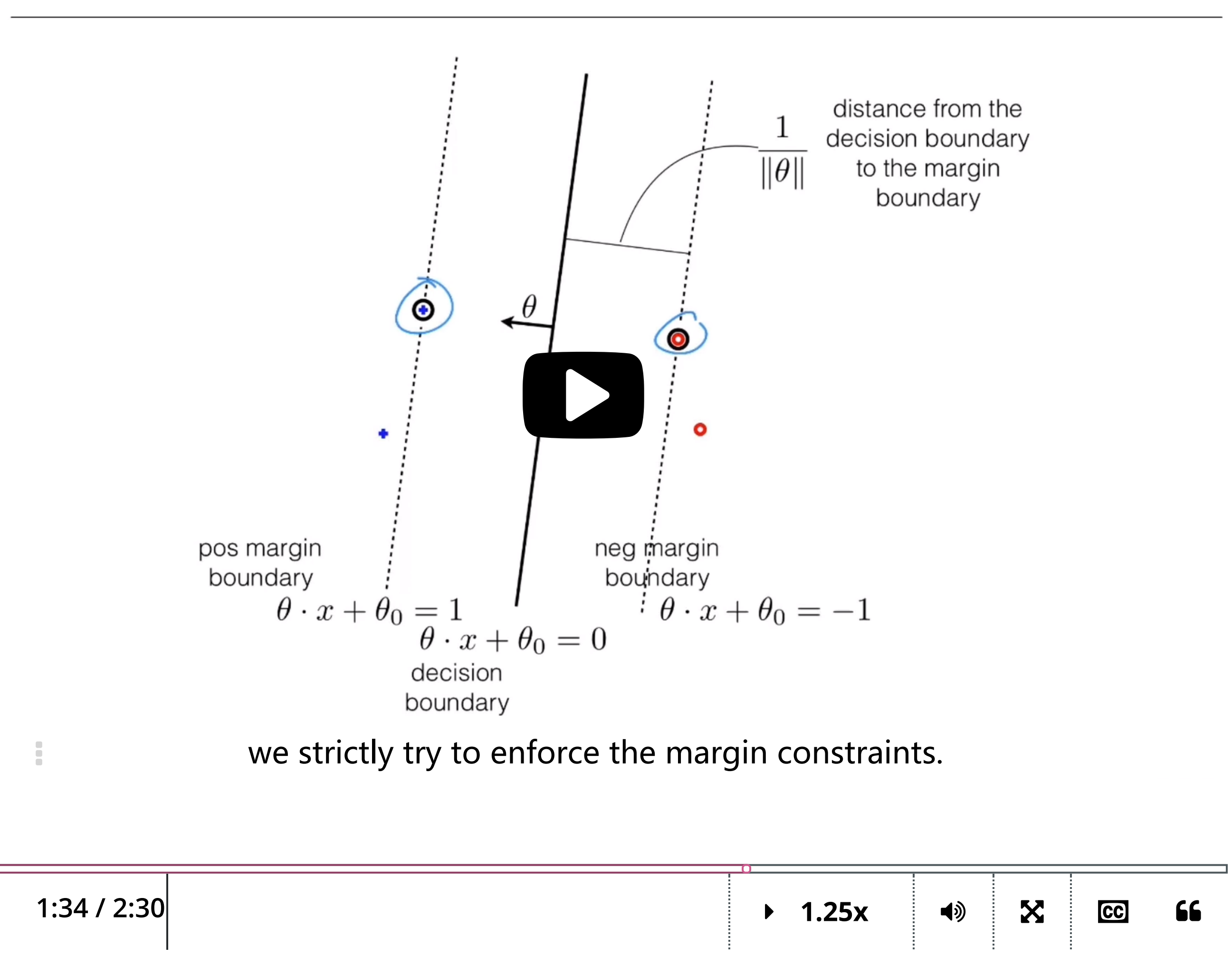
6. The Realizable Case - Quadratic program

The Realizable Case - Quadratic program



But we cannot go further because in this simple case,

we strictly try to enforce the margin constraints.

What we have seen so far is how to understand the optimization

problem corresponding to the maximum margin

linear classification, the effect of regularization as we

change the regularization parameter, how

the solution changes qualitatively

as well as in terms of generalization.

We also saw how to actually solve

the associated optimization problem using gradient descent

updates, in particular stochastic gradient descent

updates that I present to perform for such functions.

We also briefly discussed how to turn

the associated optimization problem

into a quadratic programming problem

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

The realizable case 1

1/1 point (graded)

In the realizable case, which of the following is true?

- There is exactly one (θ, θ_0) that satisfies $y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right) >= 1$ for $i = 1, \ldots n$
- There are more than one, but finite number of (θ, θ_0) that satisfy $y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right) >= 1$ for $i = 1, \ldots n$
- (selected) There are infinitely many (θ, θ_0) that satisfy $y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right) >= 1$ for $i = 1, \ldots n$ ✔

Solution:

Without any additional constraint, because θ and θ_0 are continuous, there are numerously many (θ, θ_0) that satisfy the zero-error case.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

The realizable case 2

1/1 point (graded)
Remember the objective function

$$J(\theta, \theta_0) = \frac{1}{n}\sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right)\right) + \frac{\lambda}{2}\mid\mid \theta \mid\mid^2$$

In the realizable case, we can always find (θ, θ_0) such that the sum of the hinge losses is 0. In this case, what does the objective function J reduce to?

- [ ] $\frac{1}{n}\sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right)\right)$
- [ ] $\frac{1}{n}\sum_{i=1}^{n} \text{Loss}_h\left(y^{(i)}\left(\theta \cdot x^{(i)} + \theta_0\right)\right) + \frac{\lambda}{2}\mid\mid \theta \mid\mid^2$
- [x] $\frac{1}{2}\mid\mid \theta \mid\mid^2$ ✔

Solution:

In the realizable case, we can always find a decision boundary such that the first term of $J(\theta, \theta_0)$ is 0. Thus $J(\theta, \theta_0)$ reduces to $\frac{\lambda}{2}\mid\mid \theta \mid\mid^2$. Our goal is to find θ that minimizes J anyways, so J reduces to $\frac{1}{2}\mid\mid \theta \mid\mid^2$

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Support Vectors

1/1 point (graded)
Support vectors refer to points that are exactly on the margin boundary. Which of the following is true? Choose all those apply.

- [ ] If we remove one point that is not a support vector, we will get a different θ, θ_0
- [x] If we remove all points that are support vectors, we will get a different θ, θ_0 ✔
- [ ] If we remove one point that is a support vector, we will get the same θ, θ_0
- [x] If we remove one point that is not a support vector, we will get the same θ, θ_0 ✔

✔

Solution:

Support vectors determine the exact solution θ, θ_0 that minimizes $J(\theta, \theta_0)$. Thus removing/changing all of them changes the θ, θ_0. On the other hand, any training example that is not a support vector has no influence on θ, θ_0. Thus removing/changing them does not affect θ, θ_0.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Discussion

Show Discussion

Topic: Unit 1 Linear Classifiers and Generalizations (2 weeks):Lecture 4. Linear Classification and Generalization / 6. The Realizable Case - Quadratic program