



deeplearning.ai

Regularizing your neural network

Other regularization methods

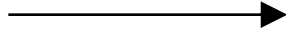
In addition to L2 regularization and dropout regularization.

Data augmentation

So data augmentation can be used as a regularization technique, in fact similar to regularization.

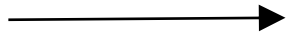
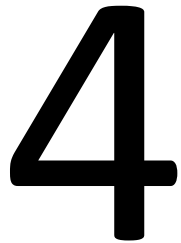
Because your training set is now a bit redundant, this isn't as good as if you had collect an additional set of brand new independent example, these extra fake training examples they don't add as much information as...

You can do this almost free other than for some confrontational costs. This can be an inexpensive way to give your algorithm more data and therefore sort of regularize it and reduce over fitting.



And by synthesizing examples like this, what you're really telling your algorithm is that something is a cat, then flipping it horizontally is still a cat.

practically, just a more subtle distortion than I'm showing here.



imposing random rotations and distortions to it.

Early stopping

Orthogonalization: this is an idea that you want to be able to think about one task at a time.

The main downside of early stopping is that this couples, these two tasks. So you no longer can work on these two problem independently because by stopping gradient descent early, you're sort of breaking whatever you're doing to optimize cost function J , because now you're not doing a great job reducing the cost function J . And then you also simultaneously trying to not over fit. So instead of using different tools to solve the two problems, you're using one tool that kind of mixes the two. And that just make the things you could try more complicated to think about.

Orthogonalization.

Momentum, RMSprop and Adam

Optimize cost function J

- Gradient, ...

Not overfit.

- Regularization, ...

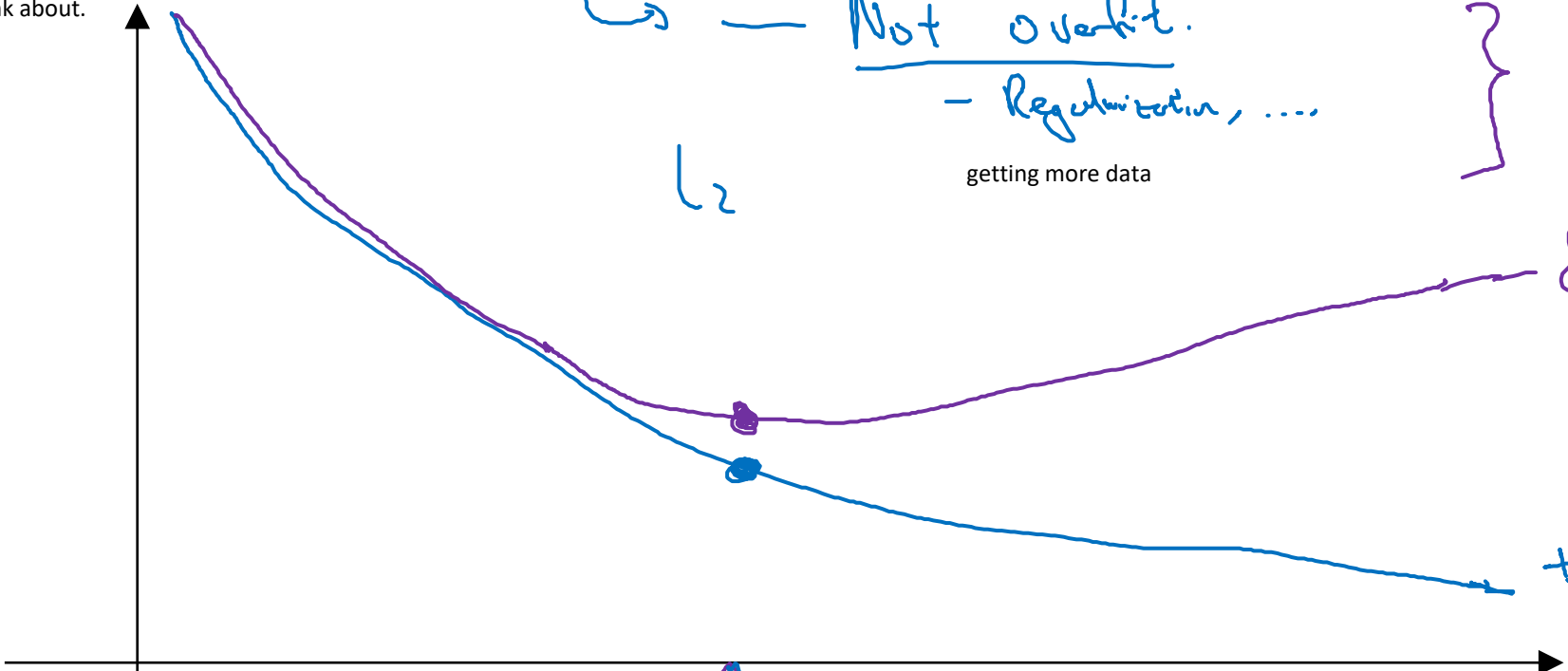
$J(w, b)$

l_2

getting more data

dev set error

train error or J



$w \approx 0$

iterations

mid-size $\|w\|_2^2$

large w