FAQ for

Lecture L05: Fitting Neurons with Gradient Descent

Follow-up Explanations 1 (From Office Hours)

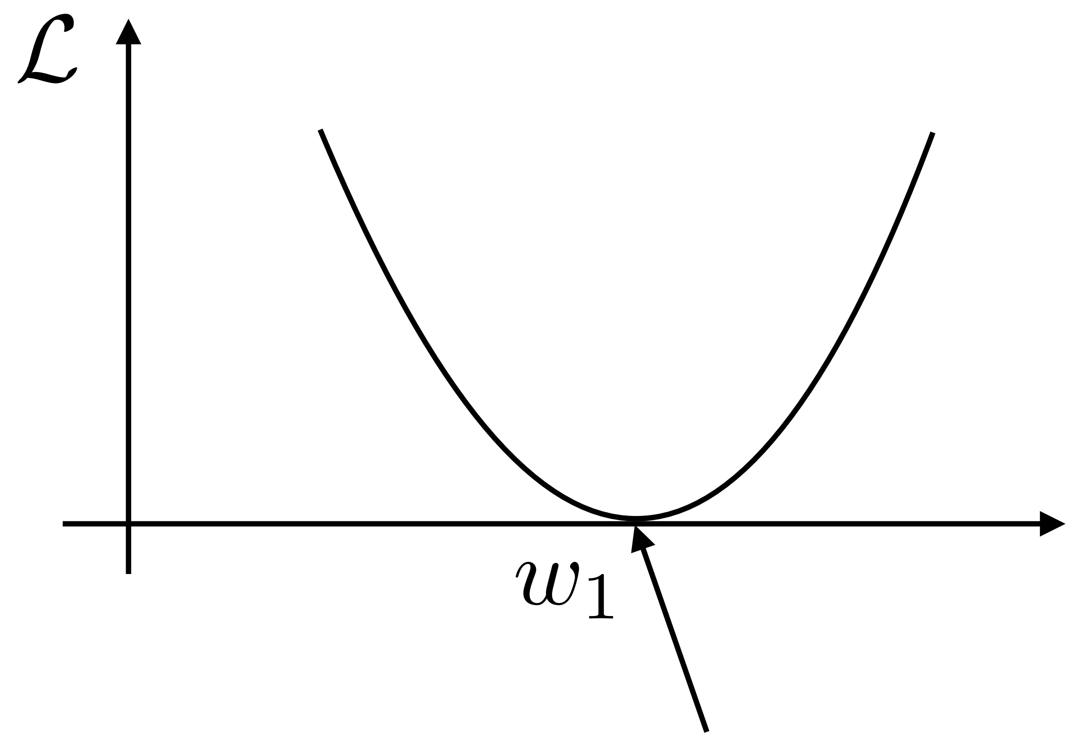
Why is the Loss convex?

Because we assumed it's the

Sum Squared Error (SSE) or Mean Squared Error (MSE)

(Not every loss is convex.)

$$SSE(y, \hat{y}) = \sum_i (y - \hat{y})^2$$



In Adaline (or linear regression), the SSE is 0 if we have a weight such that $y = \hat{y}$ for all y and $\hat{y}$. Also, note that the loss is symmetric because of the exponent "2."

Follow-up Explanations 1 (From Office Hours)

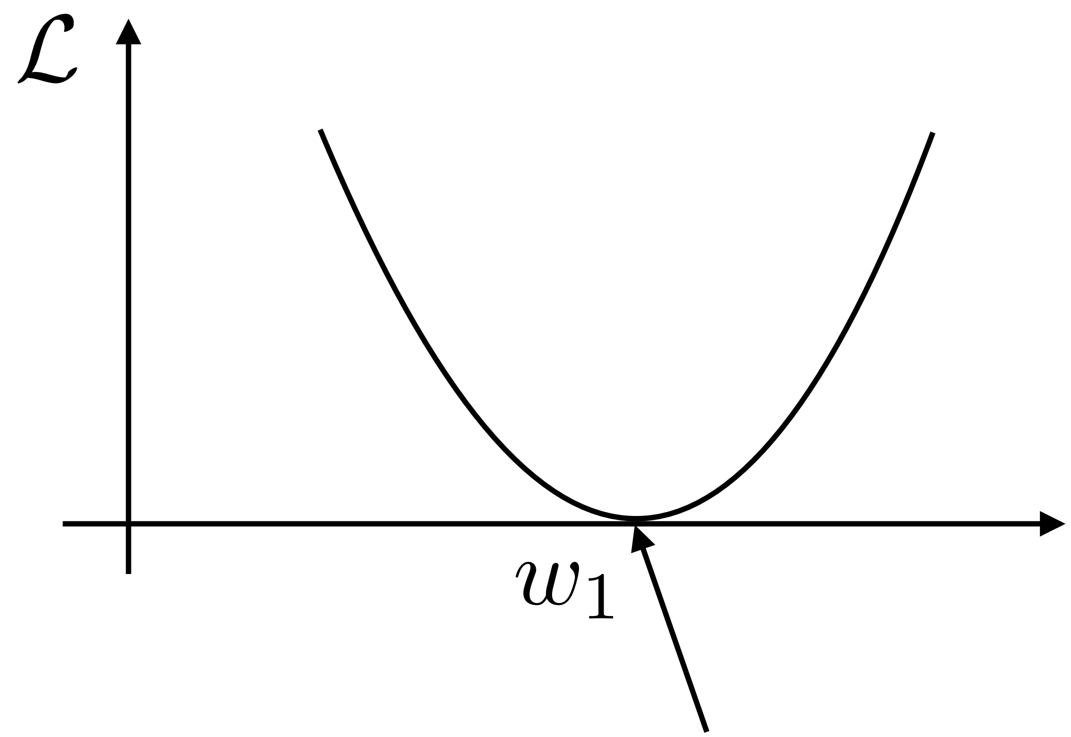
Why is the Loss convex?

Because we assumed it's the

Sum Squared Error (SSE) or Mean Squared Error (MSE)

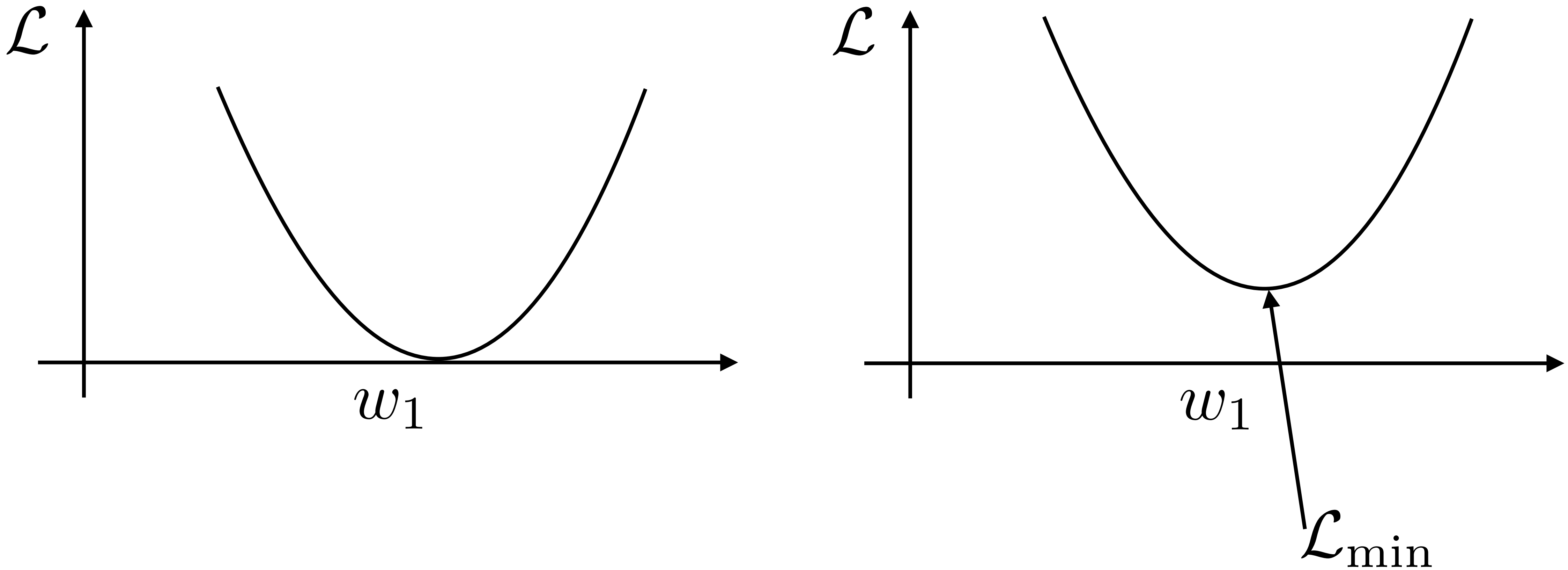
(Not every loss is convex.)

$$SSE(y, \hat{y}) = \sum_i (y - \hat{y})^2$$



In Adaline (or linear regression), the SSE is 0 if we have a weight such that $y = \hat{y}$ for all y and $\hat{y}$. Also, note that the loss is symmetric because of the exponent "2."

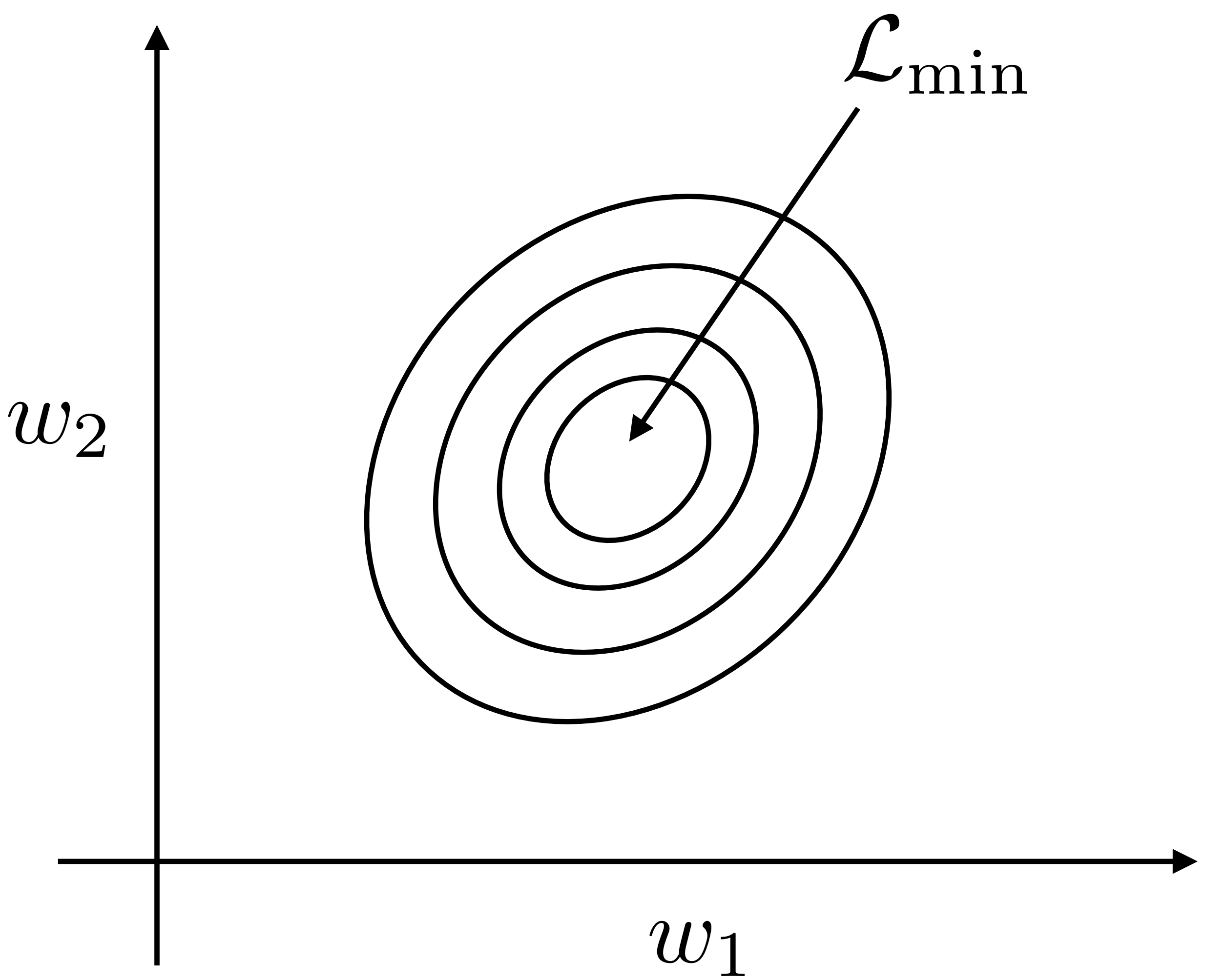
Follow-up Explanations 2 (From Office Hours)



Esp. for linear models, it is often not possible to achieve a zero loss even on the training data

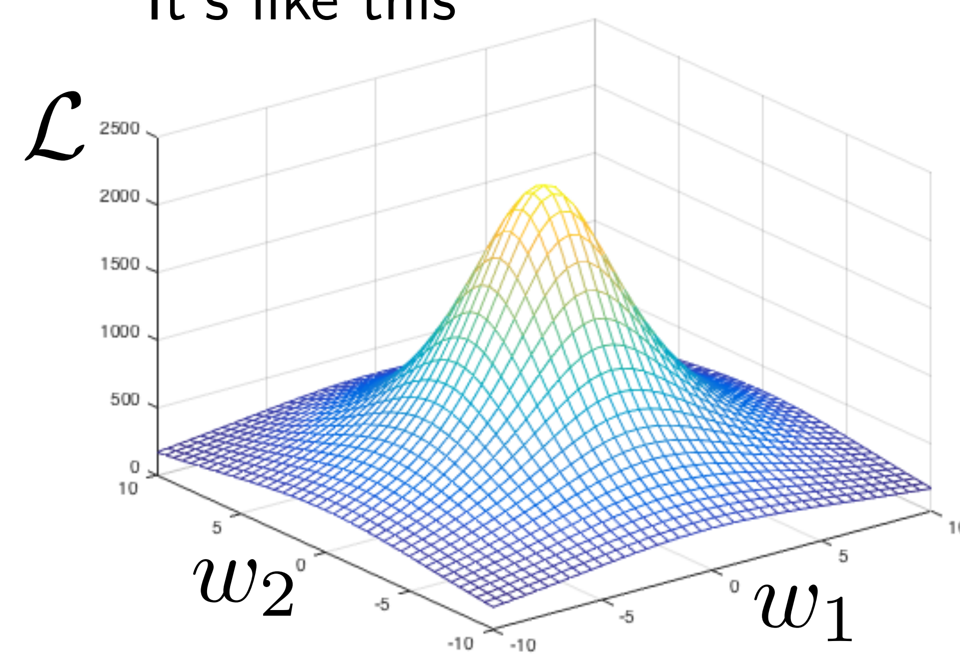
Follow-up Explanations 3 (From Office Hours)

What does this figure mean/show?

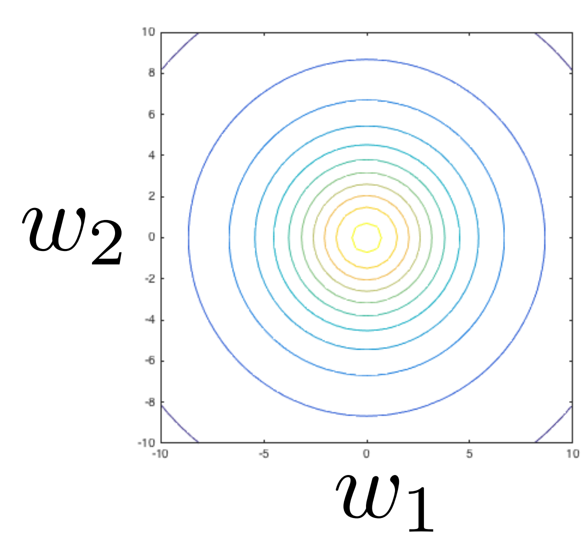


It's simply showing the loss plot (previous slide) for 2 instead of 1 weight

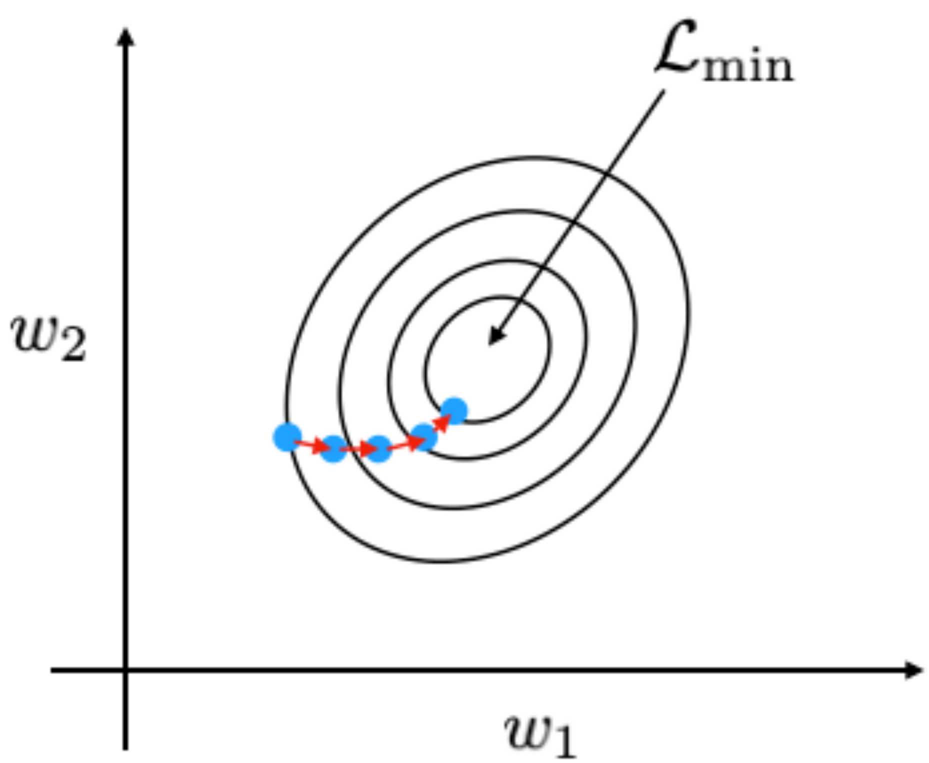
It's like this

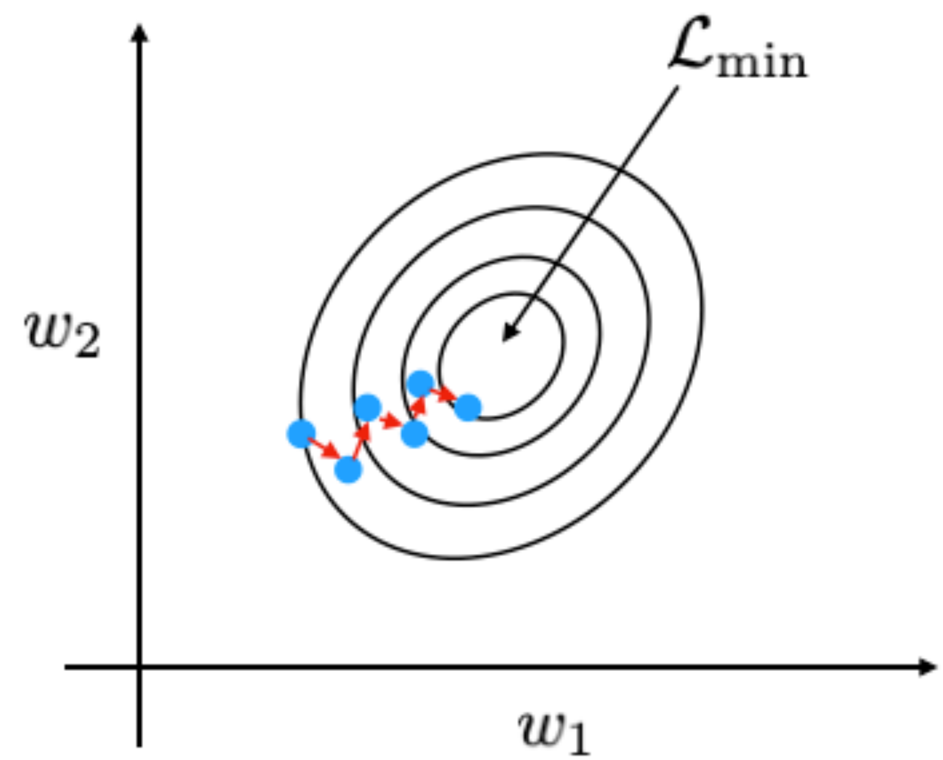


It's like this but flattened



Follow-up Explanations 4 (From Office Hours)





Why is stochastic (on-line or minibatch) noisier than batch ("whole-training-set") gradient descent?

Follow-up Explanations 5 (From Office Hours)

Why is stochastic (on-line or minibatch) noisier than batch ("whole-training-set") gradient descent?

1. Imagine you are a scientist who develops a new pharmaceutical drug.
2. You want to know its average efficiency to further improve the formula.
3. In order to know the average effectiveness, you would have to test this on all patients in the world.
4. This would be very expensive and take a long time before you get feedback! (Like "batch gradient descent").
5. Instead, you select a smaller group of patients (like a "minibatch").
6. Your estimate will be an estimate of the true average effectiveness. The larger the sample size, the better your estimate but the higher the cost; assume a certain sample size is enough such that the estimate is accurate enough that it will point you in the right direction when developing your drug...