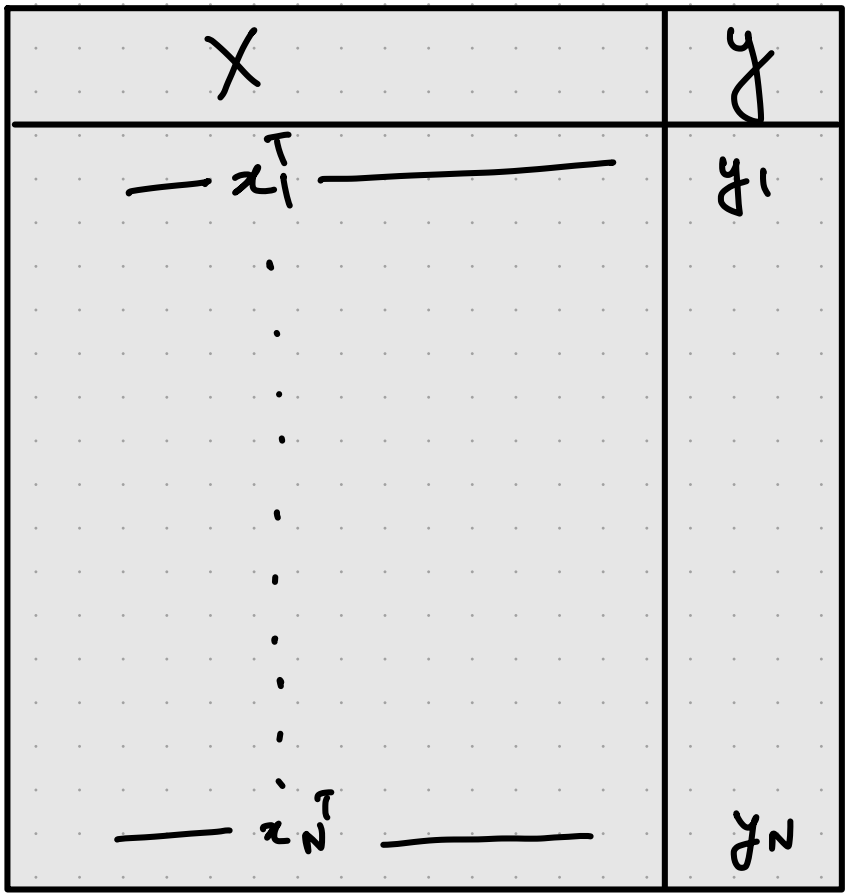
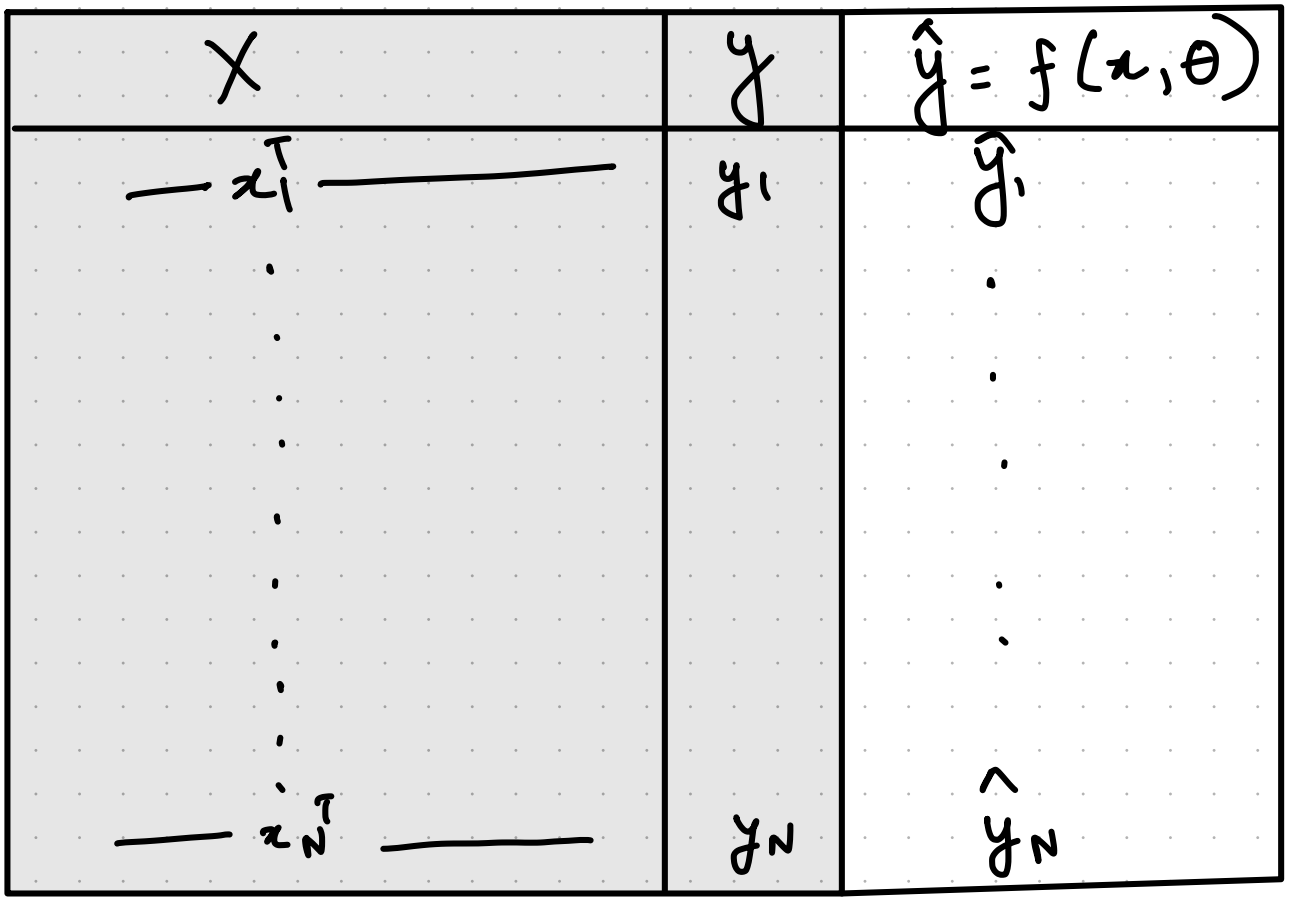
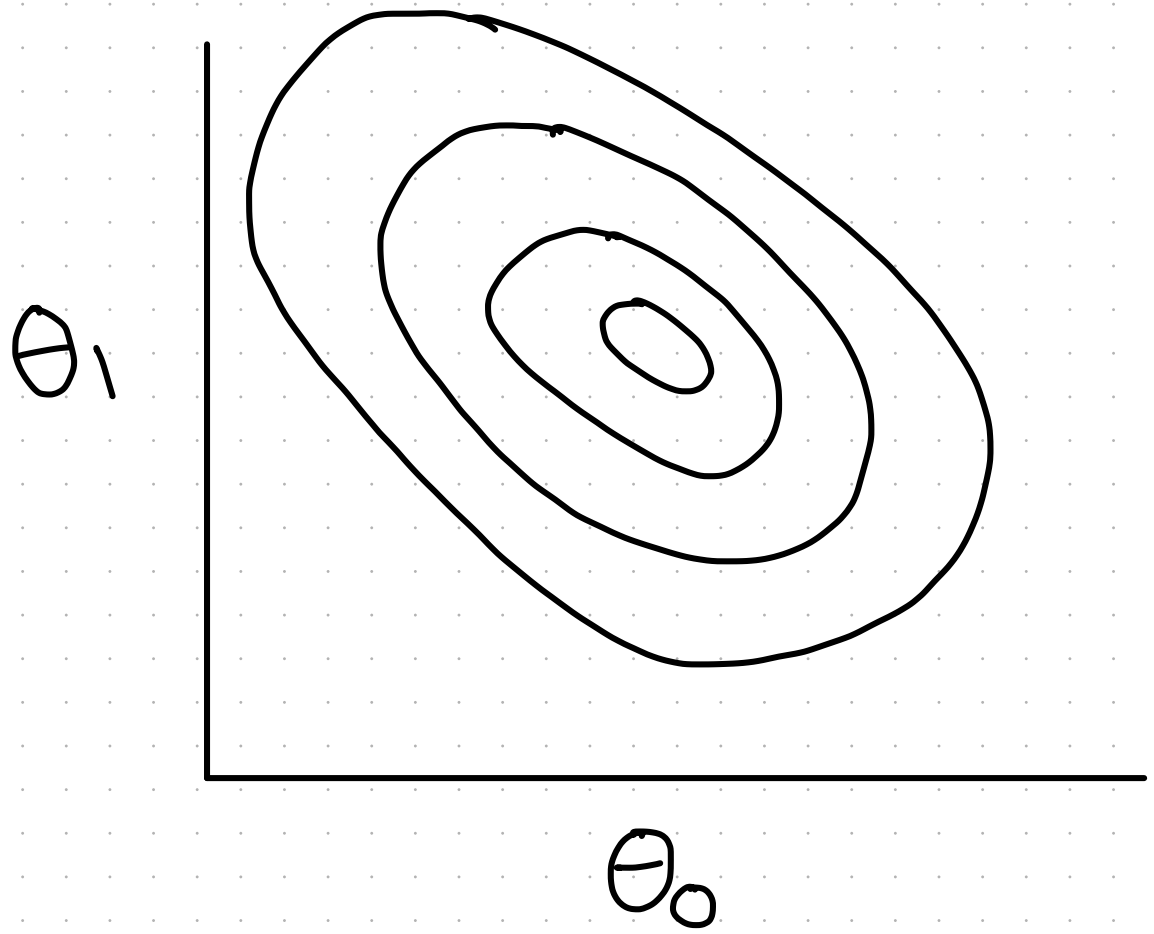
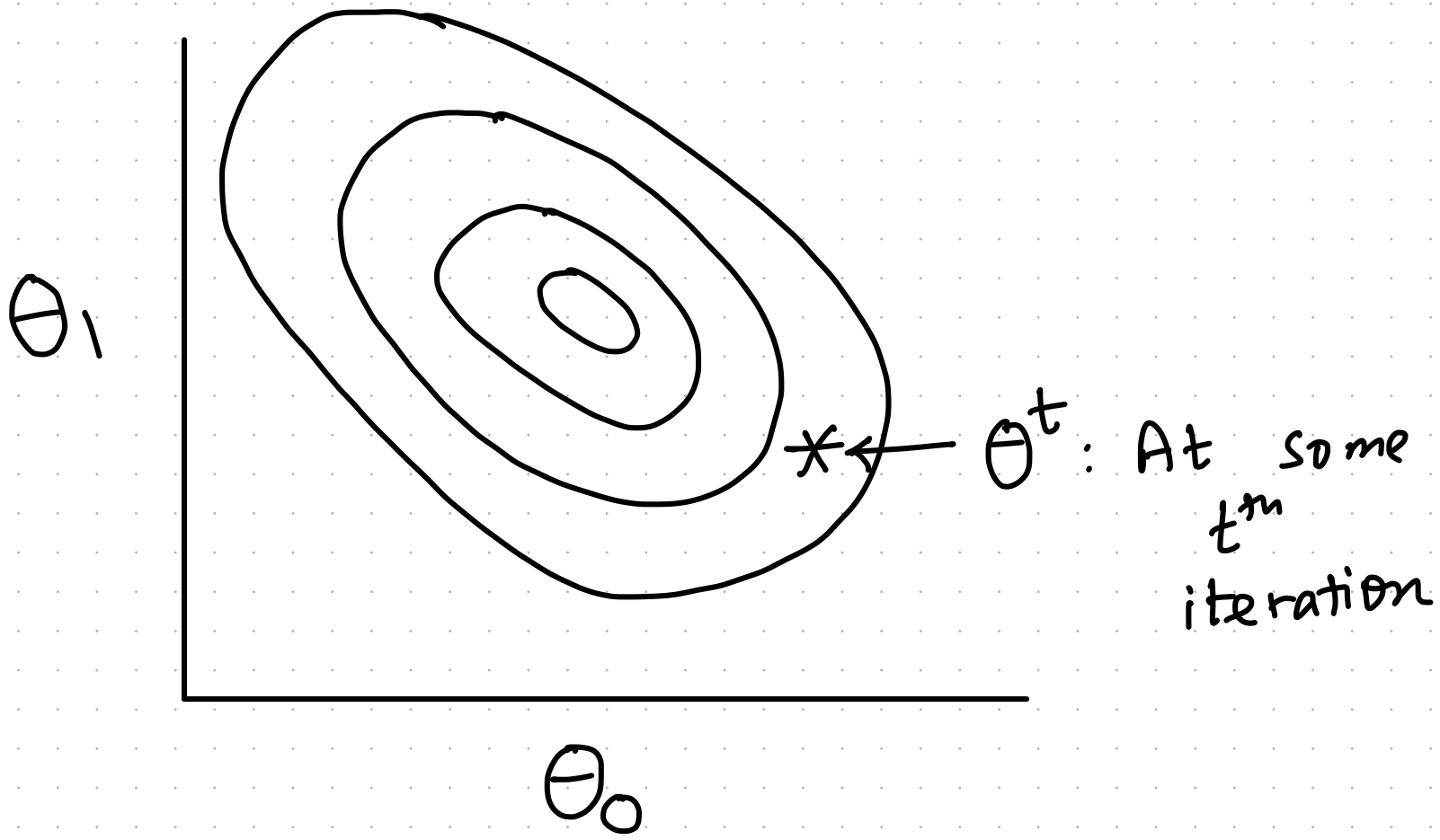

X
y
x_1^T
y_1
x_N^T
y_N

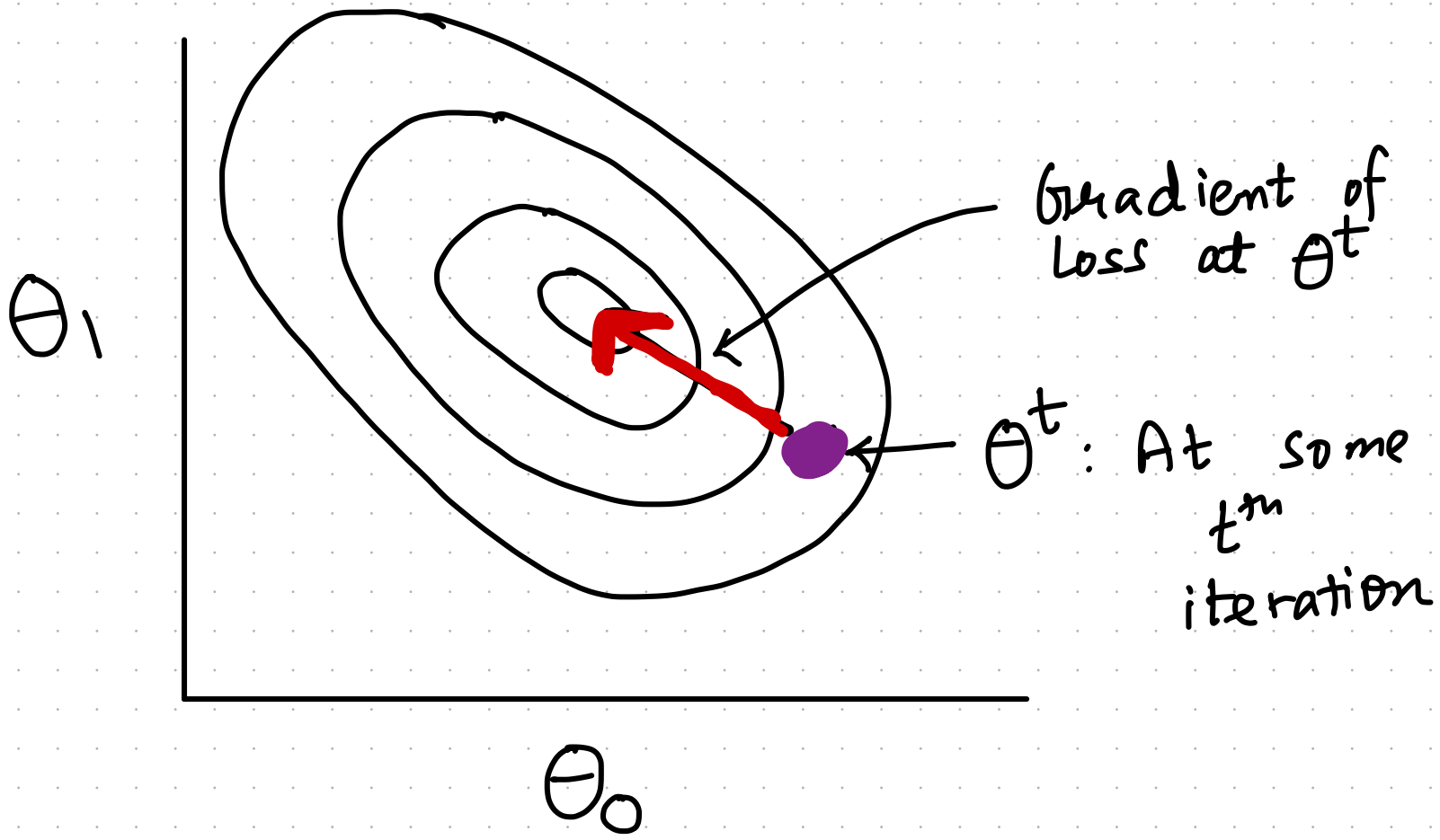

X
y
$\hat{y} = f(x, \theta)$
x_1^T
x_N^T
y_1
y_N
$\hat{y}_1$
$\hat{y}_N$



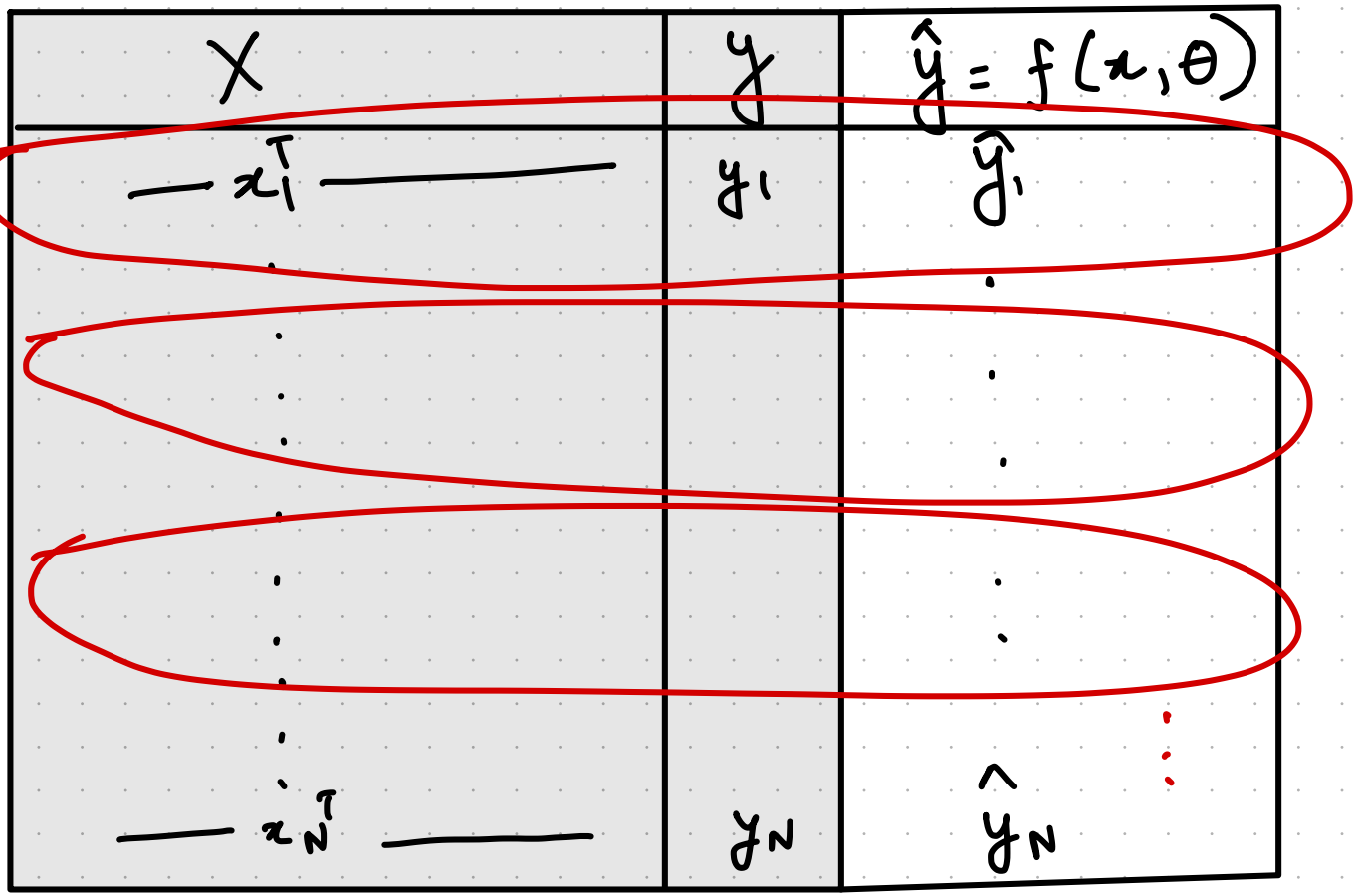
LOSS SURFACE OVER m EXAMPLES



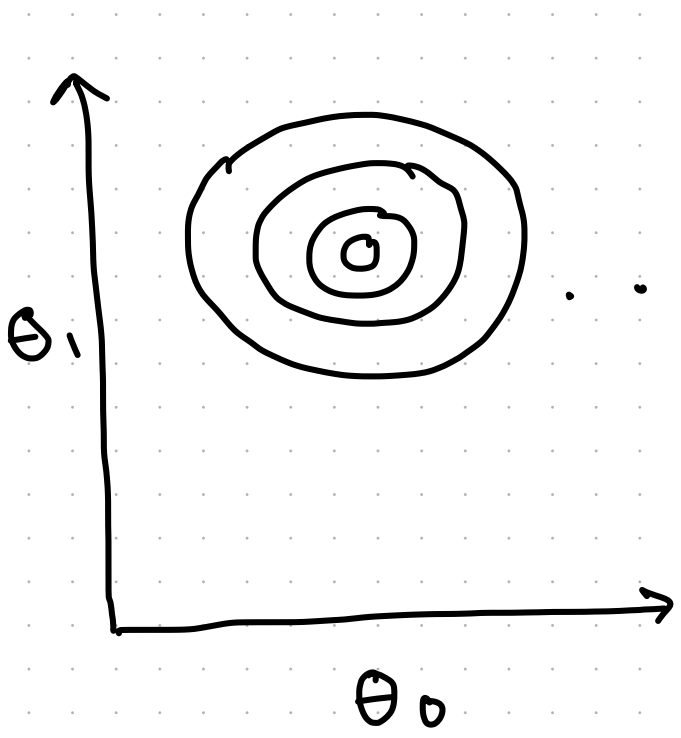
LOSS SURFACE OVER m EXAMPLES



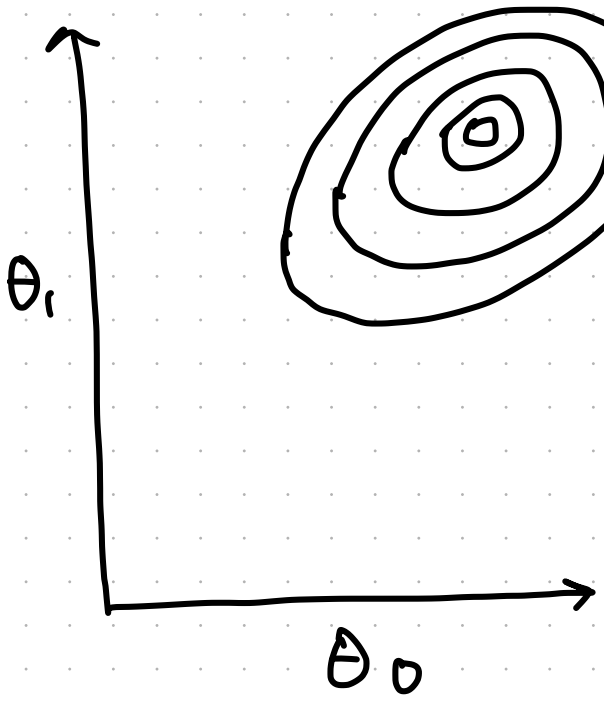
LOSS SURFACE OVER m EXAMPLES



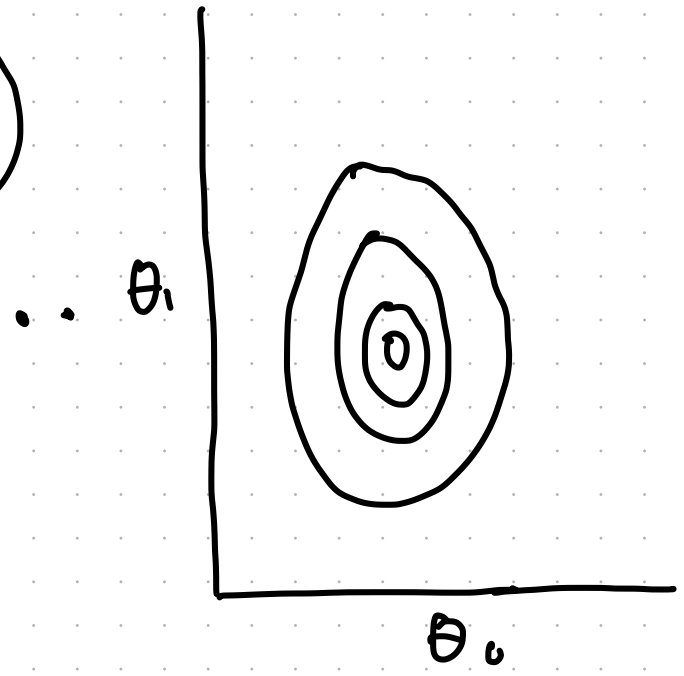
CONSIDER Individual data points to compute loss



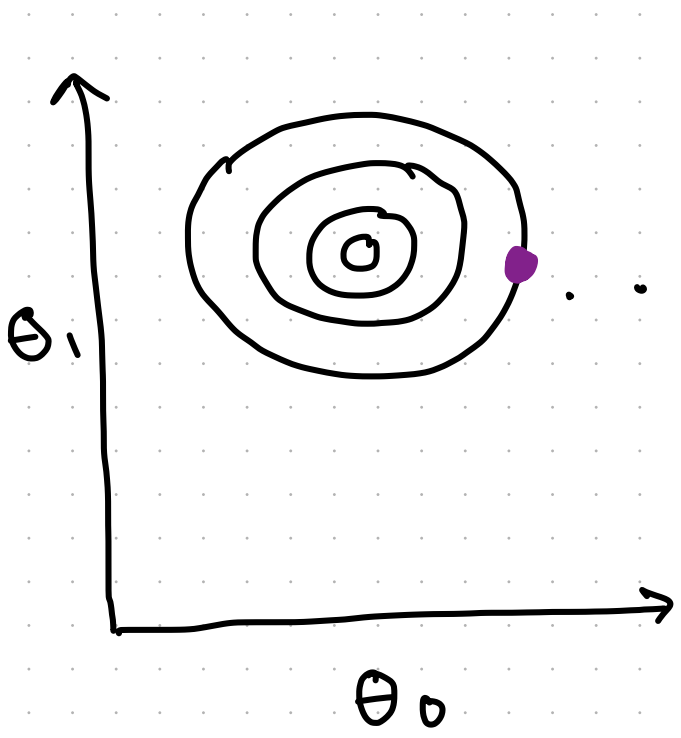
$loss(y_1, \hat{y}_1)$



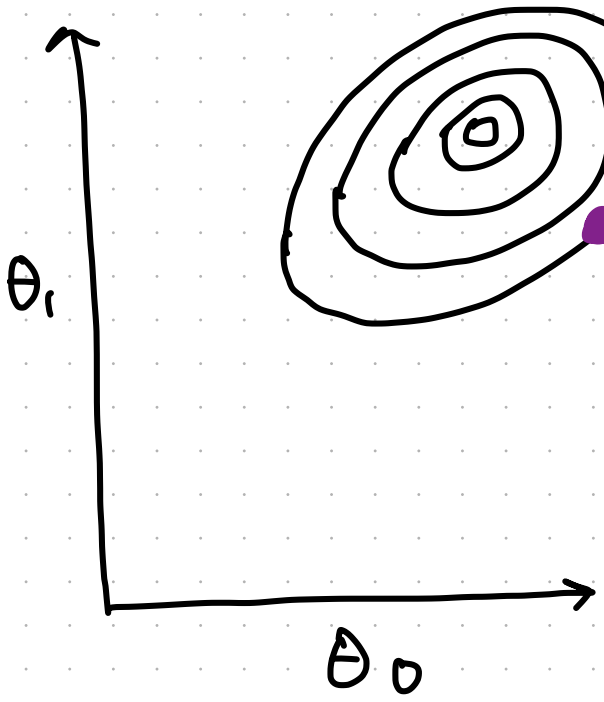
$loss(y_i, \hat{y}_i)$



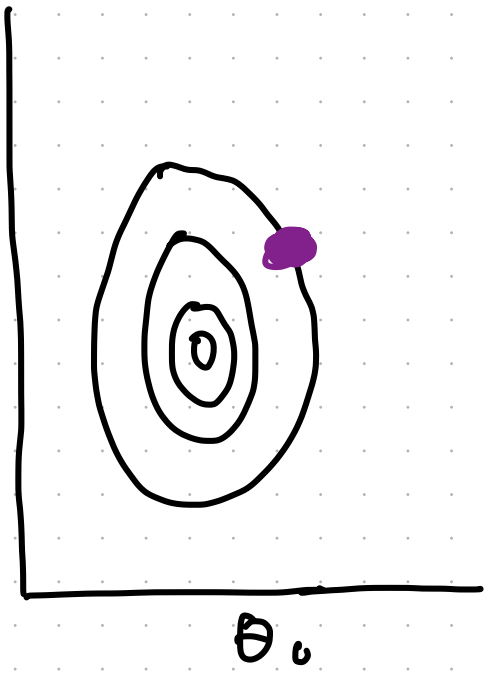
$loss(y_N, \hat{y}_N)$



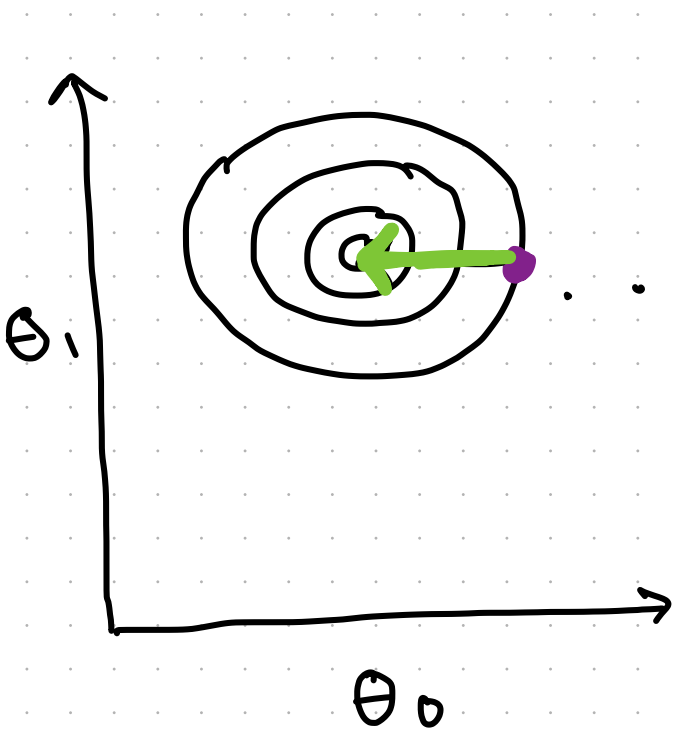
$loss(y_1, \hat{y}_1)$



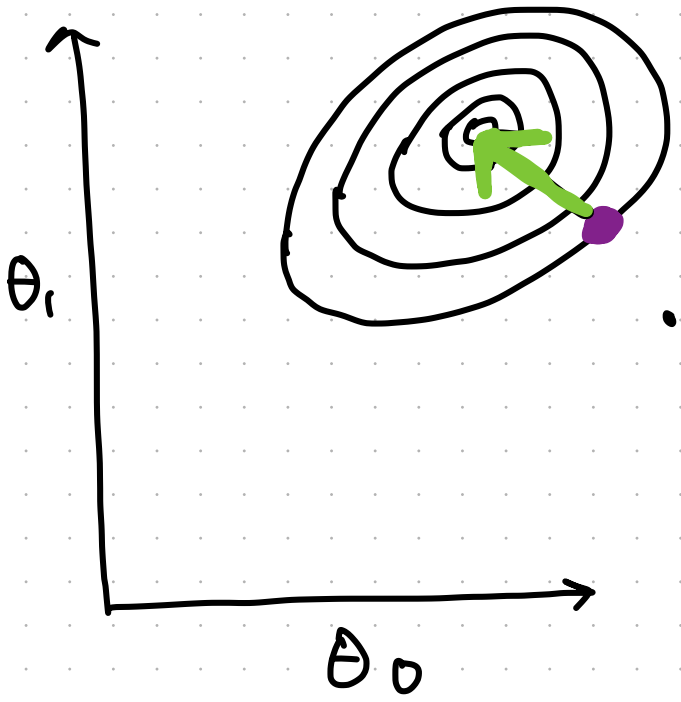
$loss(y_i, \hat{y}_i)$



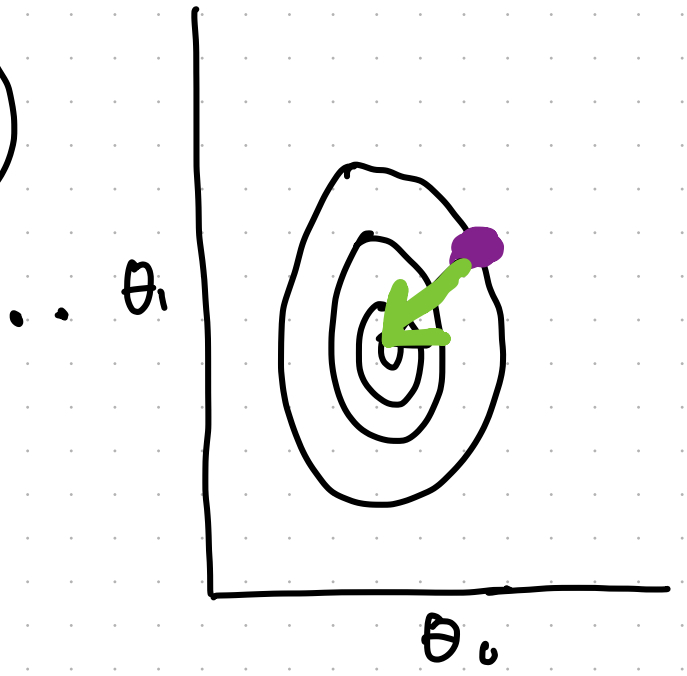
$loss(y_N, \hat{y}_N)$



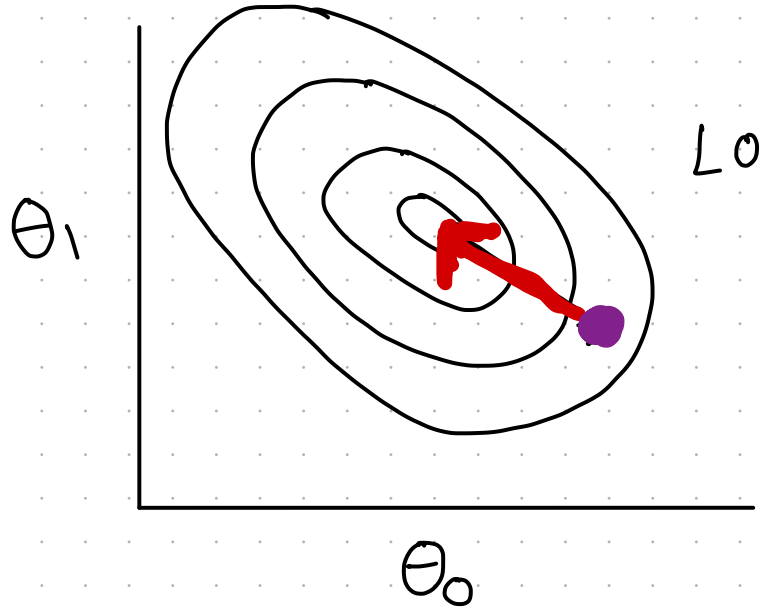
$loss(y_1, \hat{y}_1)$



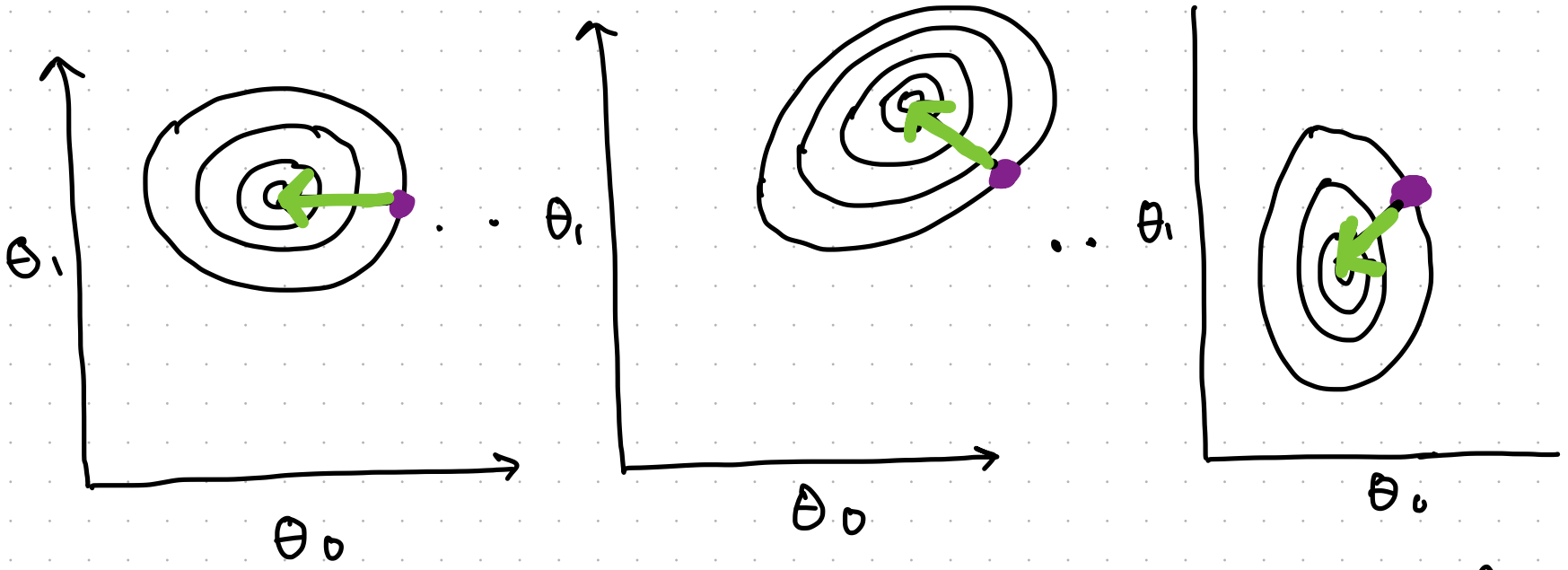
$loss(y_i, \hat{y}_i)$



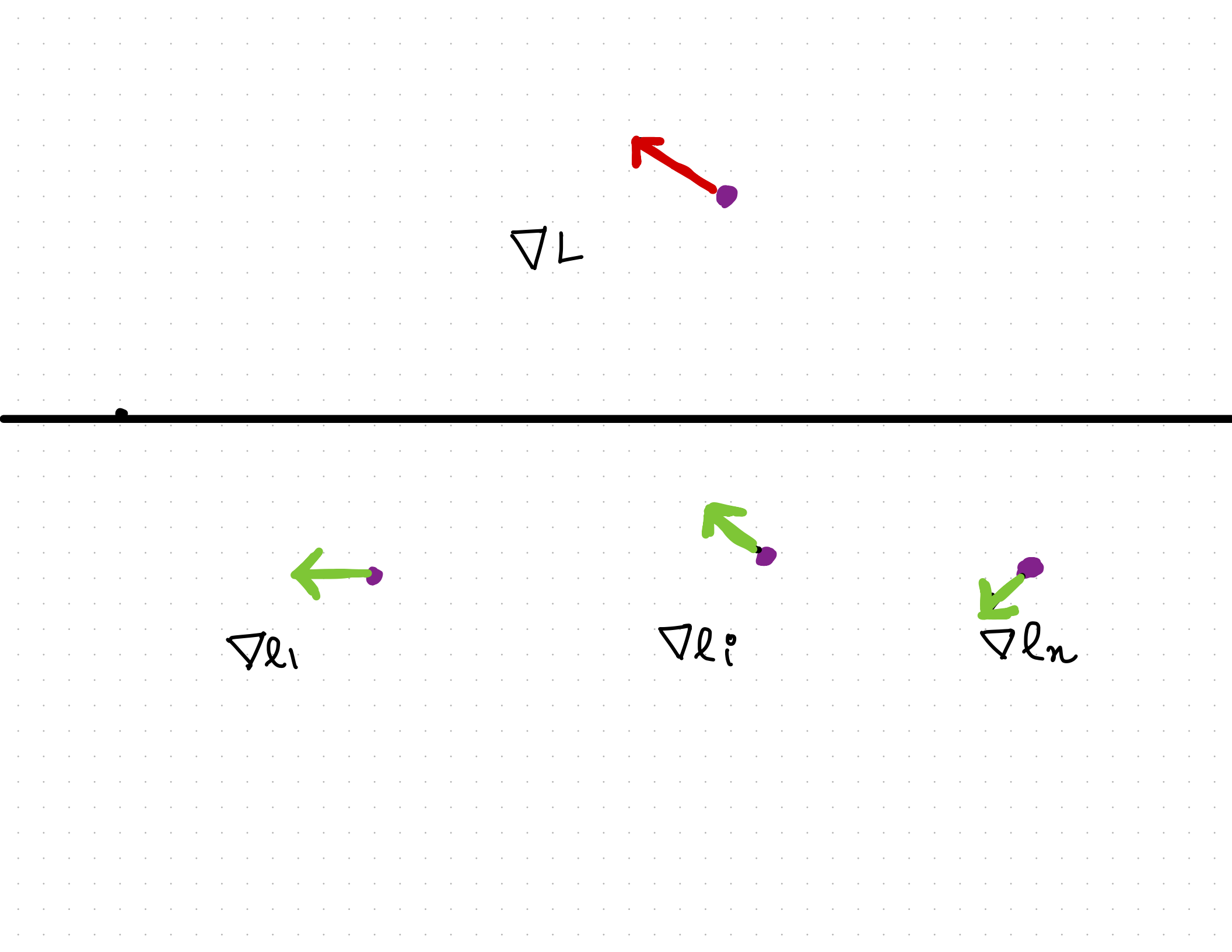
$loss(y_N, \hat{y}_N)$

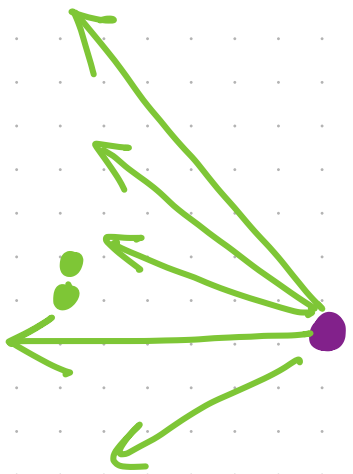


LOSS SURFACE OVER 'N' EXAMPLES

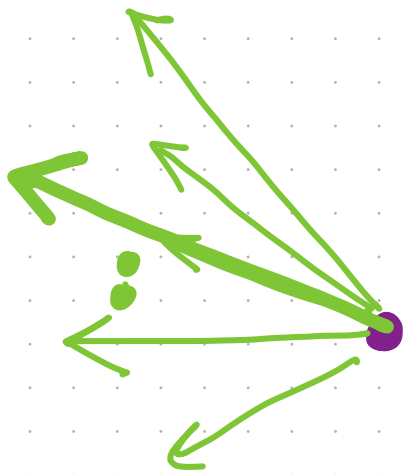


$loss(y_1, \hat{y}_1)$ $loss(y_i, \hat{y}_i)$ $loss(y_N, \hat{y}_N)$


∇L
∇l1
∇li
∇ln

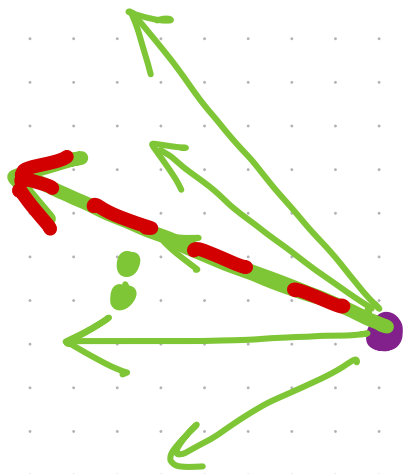


Gradients for losses wrt different points



— Gradients for losses wrt different points

— Expectation over individual gradients



Gradients for losses wrt different points

Expectation over individual gradients

Gradient wrt. whole data