

X

$$x_1^T \quad \dots$$

y

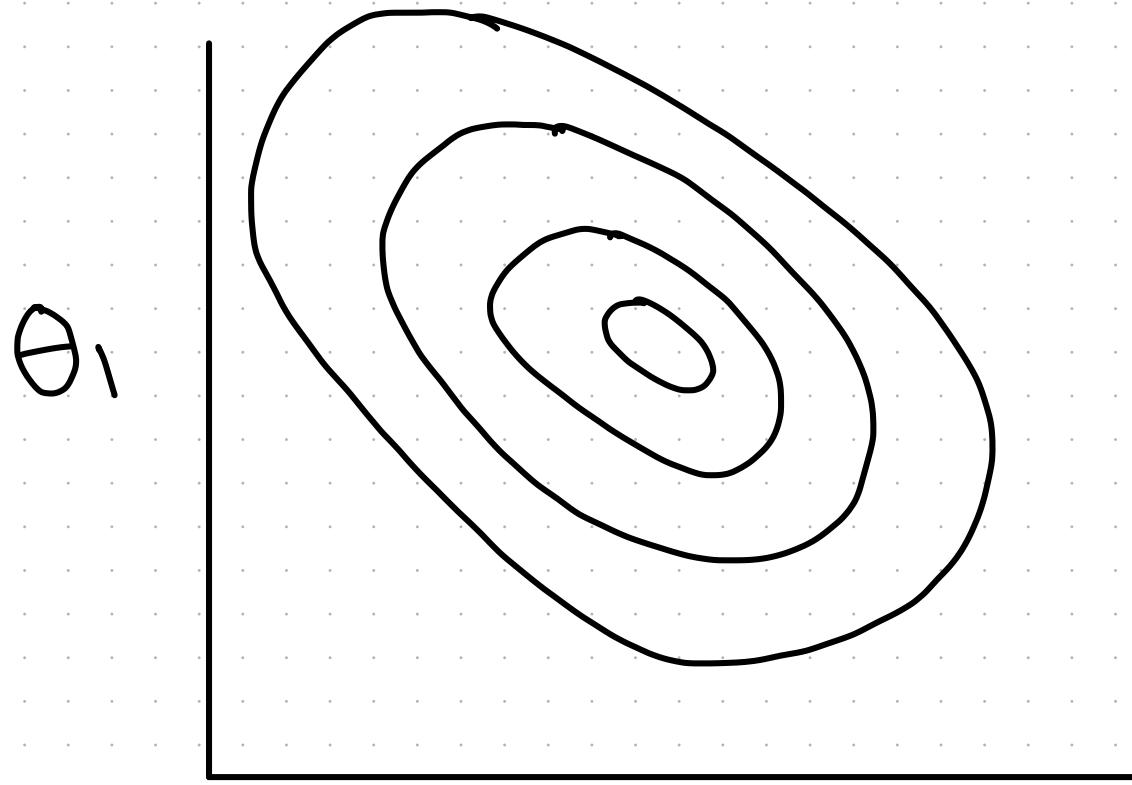
$$y_1$$

$$\dots$$

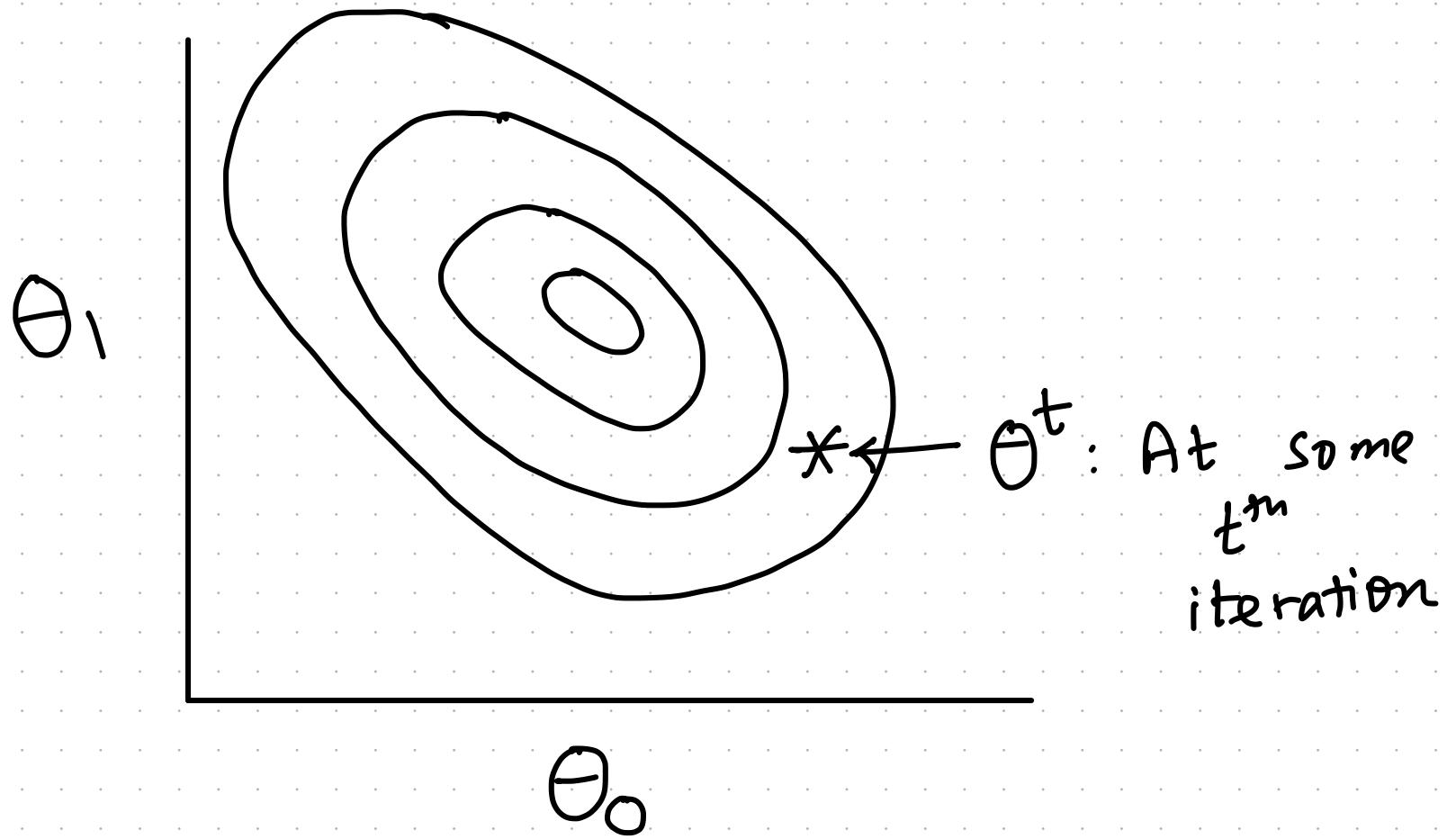
$$x_n^T \quad \dots$$

$$y_n$$

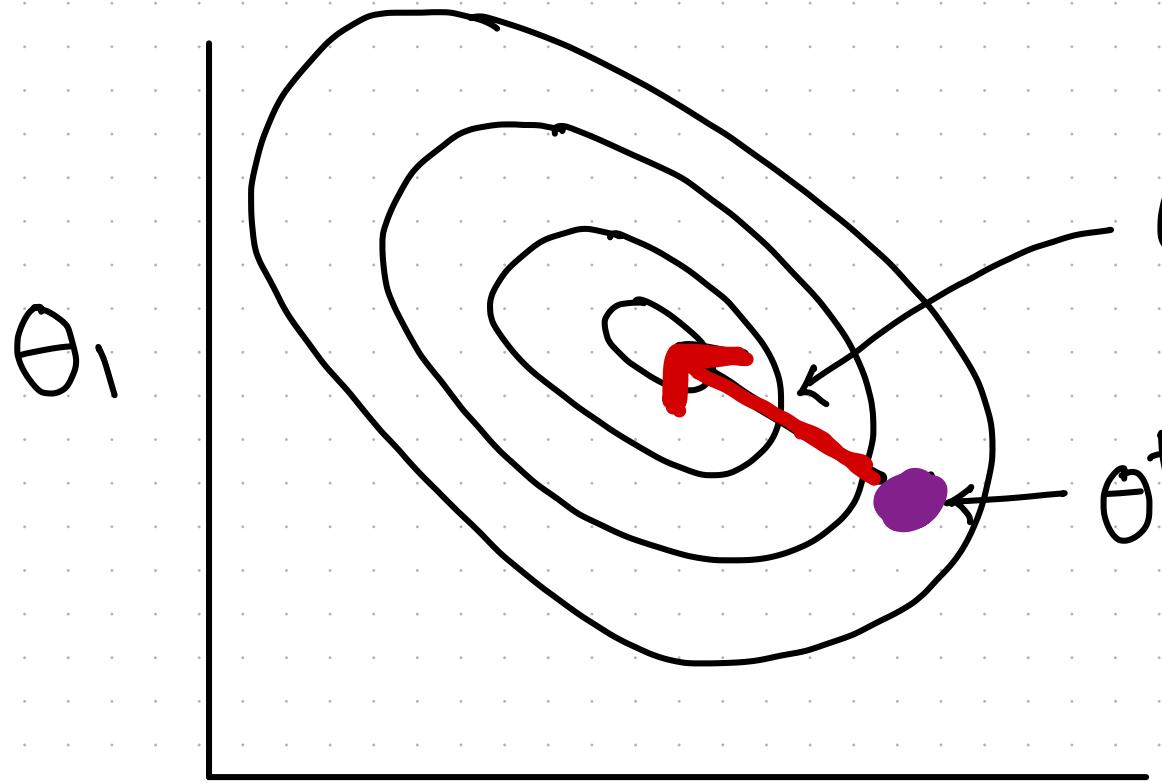
X	y	$\hat{y} = f(x, \theta)$
x_1^T	y_1	\hat{y}_1
.	.	.
.	.	.
.	.	.
x_n^T	y_n	\hat{y}_n



LOSS SURFACE OVER
 $6N'$ EXAMPLES



LOSS SURFACE OVER
 n' EXAMPLES



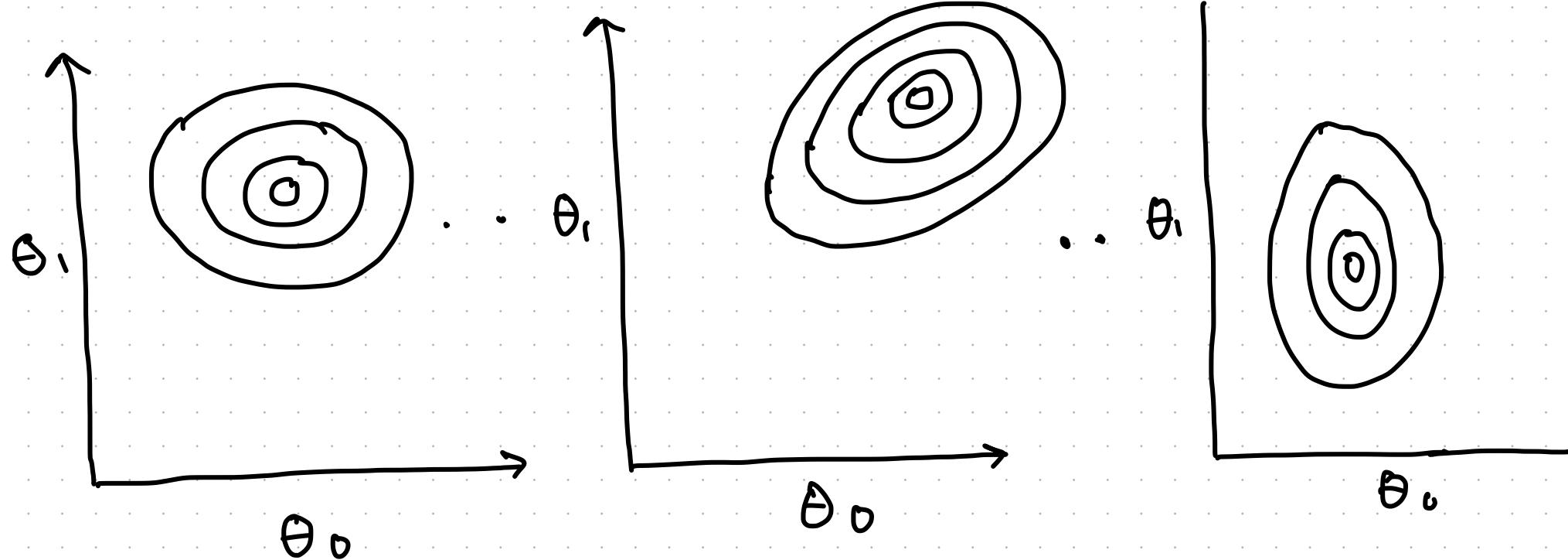
Gradient of
loss at θ^t

θ^t : At some
 t^m
iteration

LOSS SURFACE OVER
 $6N'$ EXAMPLES

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

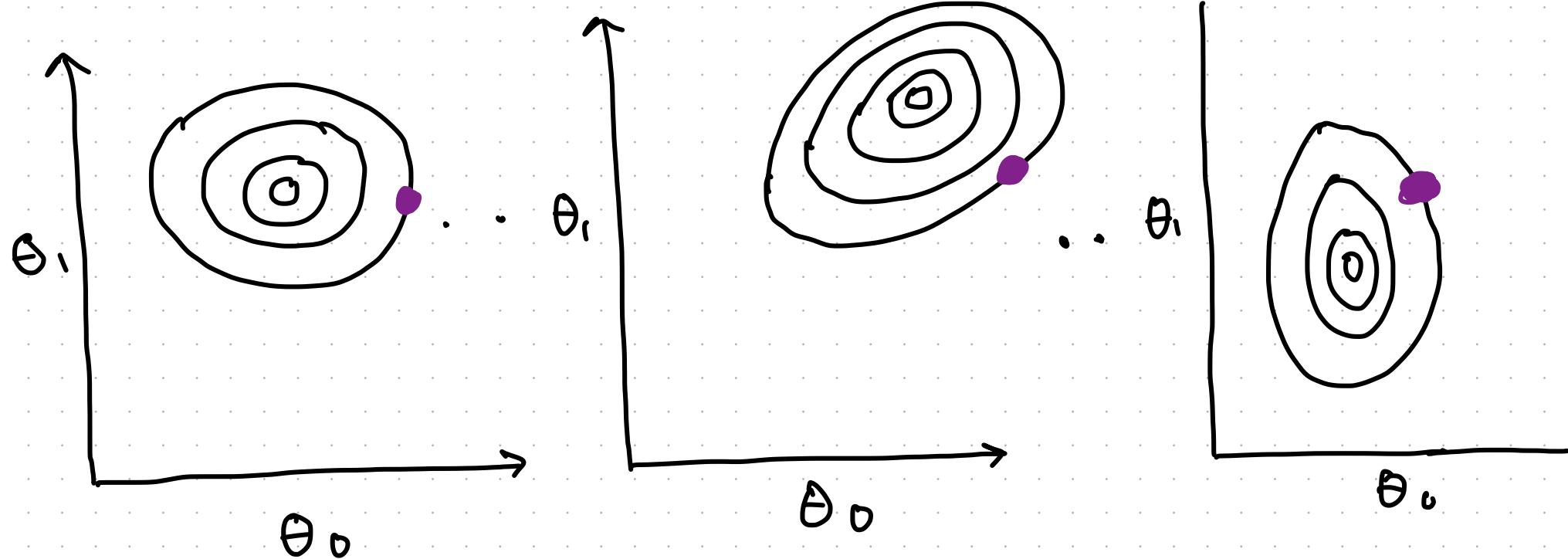
CONSIDER
Individual
data points
to compute
Loss



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

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

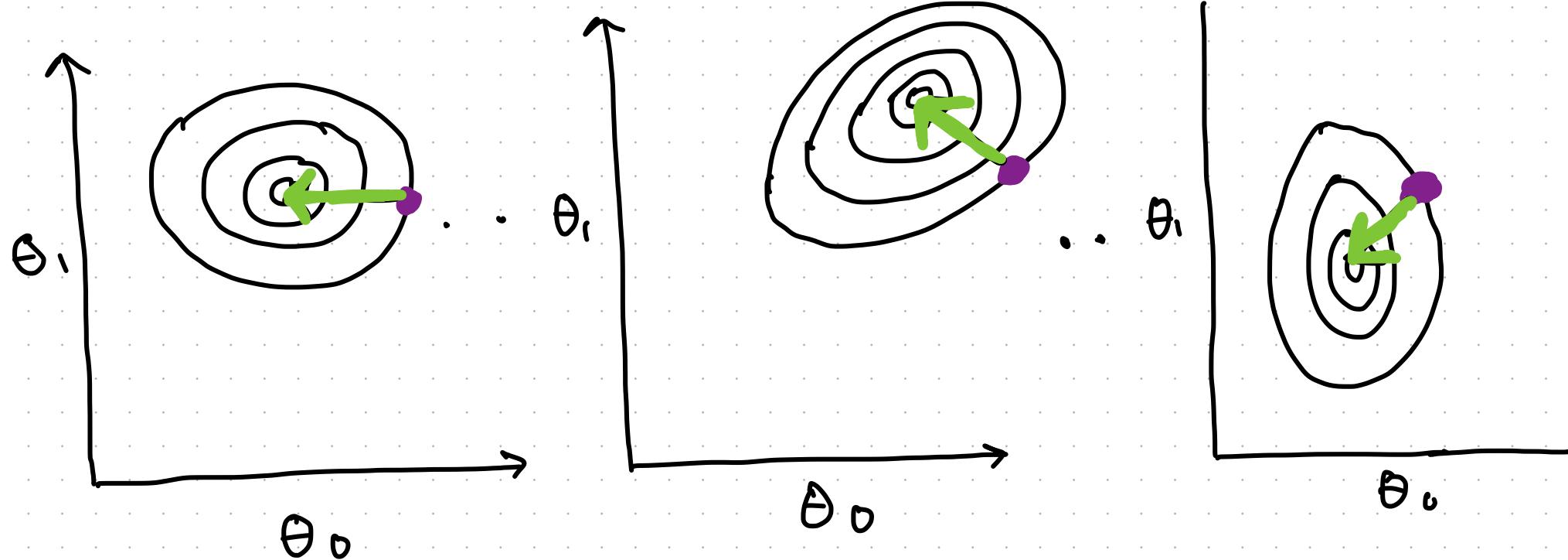
$\text{loss}(y_N, \hat{y}_N)$



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

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

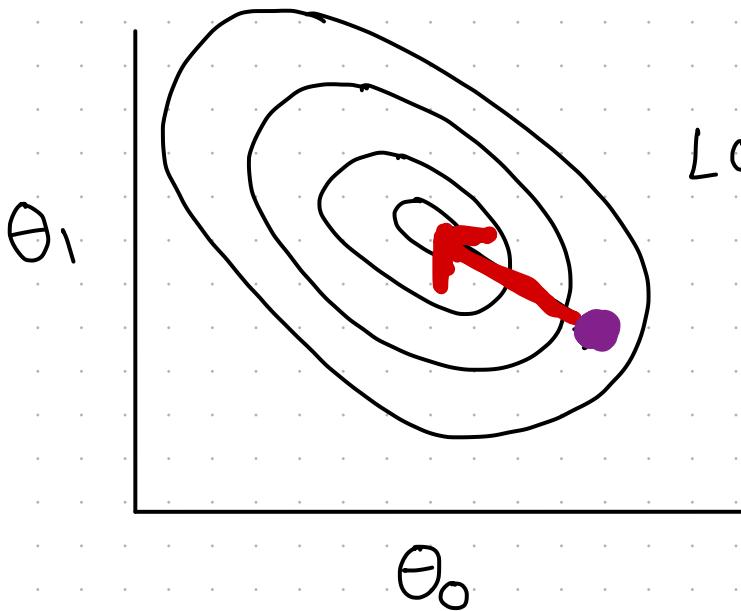
$\text{loss}(y_N, \hat{y}_N)$



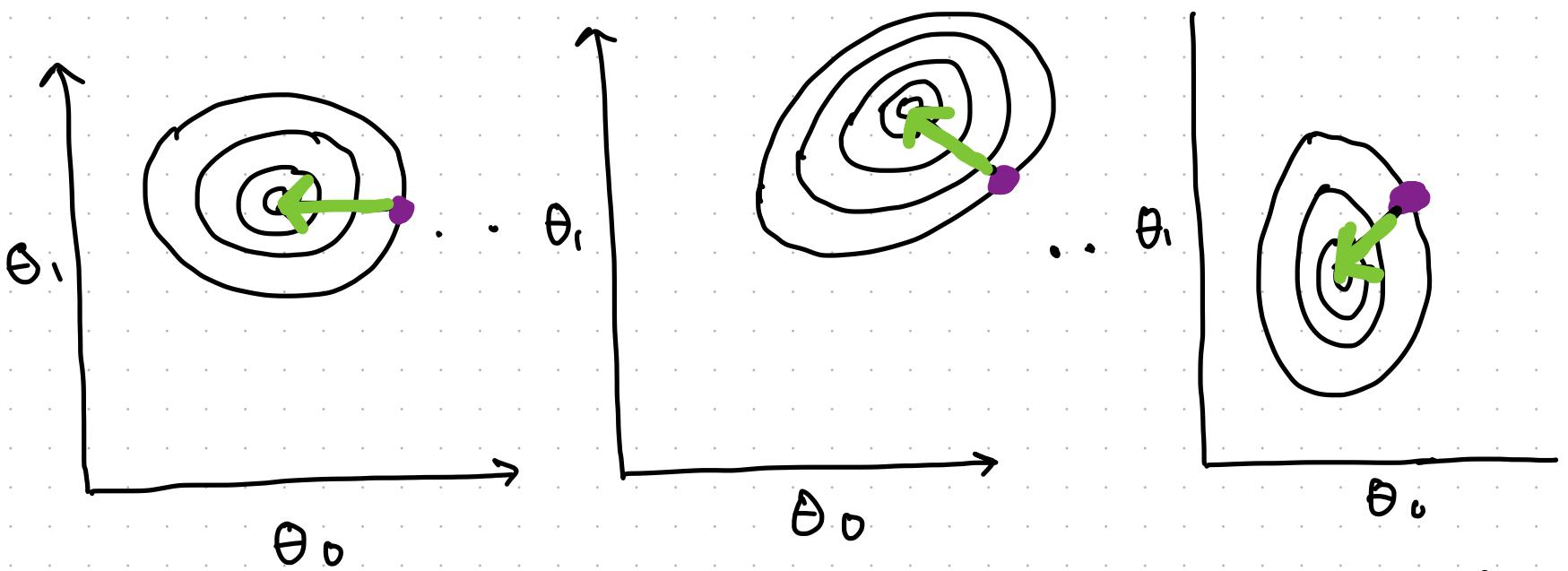
$\text{loss}(y_1, \hat{y}_1)$

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

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



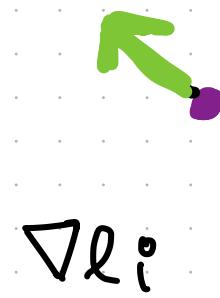
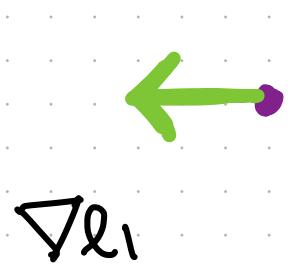
LOSS SURFACE OVER
 $6N^i$ EXAMPLES



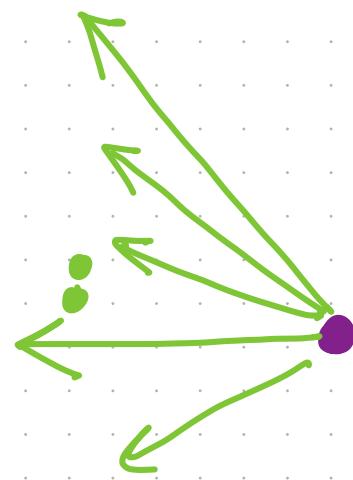
$\text{loss}(y_1, \hat{y}_1)$

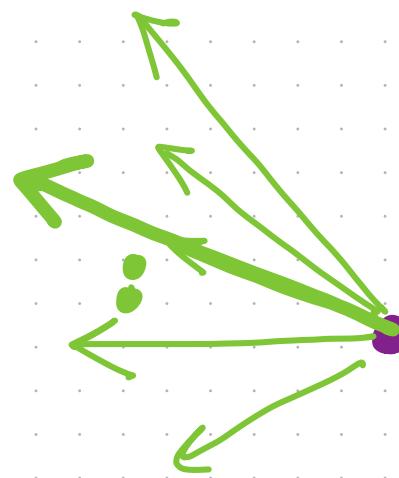
$\text{loss}(y_i, \hat{y}_i)$

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



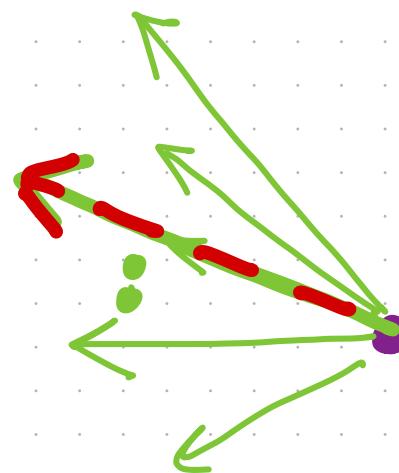
— Gradients for
losses wrt
different
points





— Gradients for losses wrt different points

— Expectation over individual gradients



- Gradients for losses wrt different points
- Expectation over individual gradients
- Gradient wrt. whole data