

# Basic Machine Learning, Gradient Descent

Parameters  $\bar{w}$  to optimize

Labeled data  $(\bar{x}^{(i)}, y^{(i)})_{i=1}^D$   $D$  points  
input label

Suppose 10,000 feats.  $\bar{w} \in \mathbb{R}^{10,000}$ . Searching for an optimal  $\bar{w}$

Optimization problem: formulate training objective, which is (for us) linear over examples, optimize

objective:  $\sum_{i=1}^D \text{loss}(\bar{x}^{(i)}, y^{(i)}, \bar{w})$

Stochastic gradient descent

for  $t$  up to num epochs:

for  $i$  up to  $D$ :

sample  $j \sim [1, D]$

$\bar{w} \leftarrow \bar{w} - \alpha \frac{\partial}{\partial \bar{w}} \text{loss}(\bar{x}^{(j)}, y^{(j)}, \bar{w})$   
step size

gradient of the loss  
w.r.t.  $\bar{w}$   
points towards  $\bar{w}$  that give  
higher loss