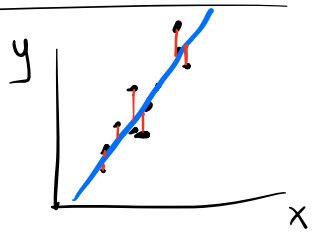
ML1 - Optimization with Gradient Descent

Linear regression

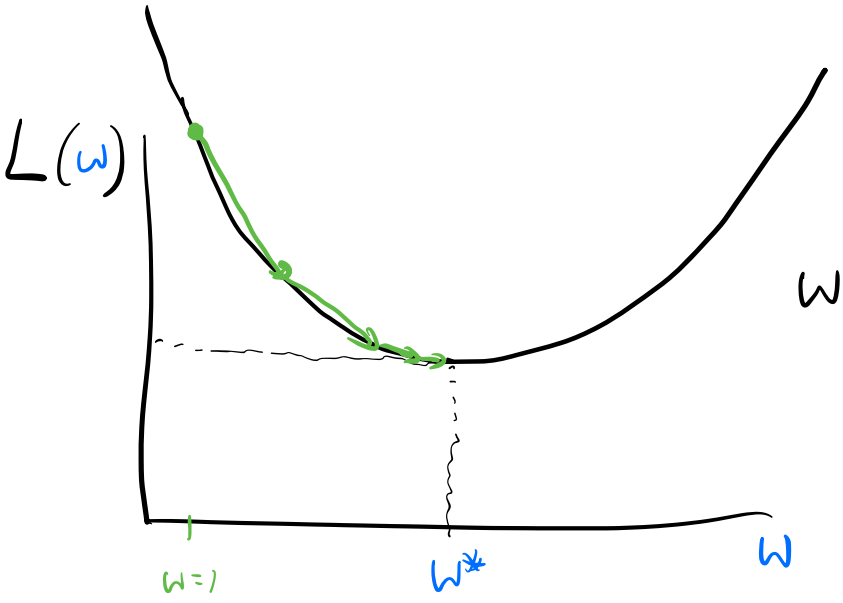


Data: $\{(x_i, y_i) : 1 \le i \le n\}$

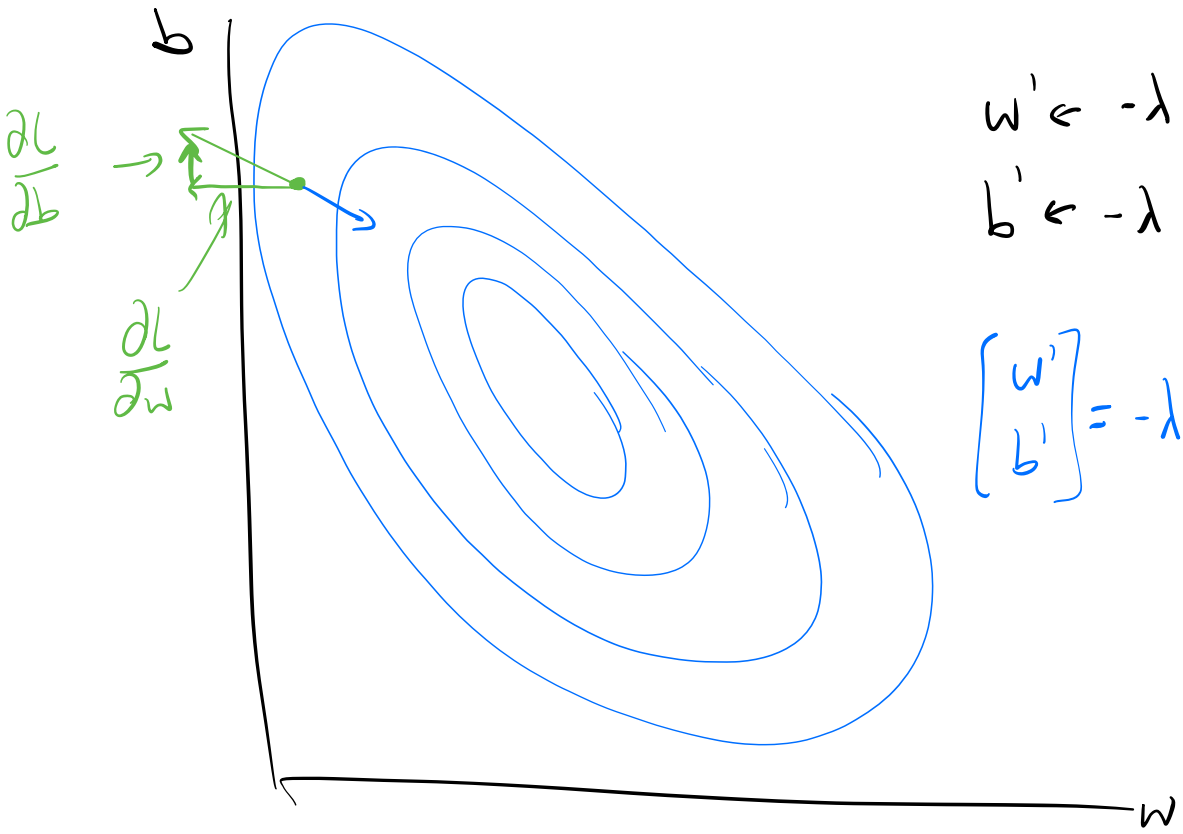
Model: $\hat{y} = wx + b$ (univariate linear regression)

Loss: $\frac{1}{2n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ (mean squared error)

$$L(w,b) = \frac{1}{2n} \sum \left(y_i - (w x_i + b) \right)^2$$



$$w' \leftarrow -\lambda \frac{d}{dw} L(w)$$



$$w' \leftarrow -\lambda \frac{\partial L}{\partial w}$$

$$b' \leftarrow -\lambda \frac{\partial L}{\partial b}$$

$$\begin{bmatrix} w' \\ b' \end{bmatrix} = -\lambda \begin{bmatrix} \frac{\partial L}{\partial w} \\ \frac{\partial L}{\partial b} \end{bmatrix}$$

$$\frac{\partial}{\partial \hat{y}_i}(y_i - \hat{y}_i)^2 = y_i - 2\hat{y}_i y_i + \hat{y}_i^2$$

$$2\hat{y}_i - 2y_i$$

$$2(\hat{y}_i - y_i)$$

$$\frac{\partial}{\partial w}(wx_i + b) = \boxed{x_i}$$

$$\frac{1}{2n}\sum x_i(\hat{y}_i - y_i)$$

Gradient Descent Algorithm:

$w, b \leftarrow w_{init}, b_{init}$

While True:

$\hat{y} \leftarrow model(x, w, b)$

$loss \leftarrow loss(y, \hat{y})$

$dw \leftarrow grad\_wrt\_w(\hat{y}, x, y)$

$db \leftarrow grad\_wrt\_b(\hat{y}, x, y)$

if $norm\begin{bmatrix} dw \\ db \end{bmatrix} < \varepsilon$:

return w, b

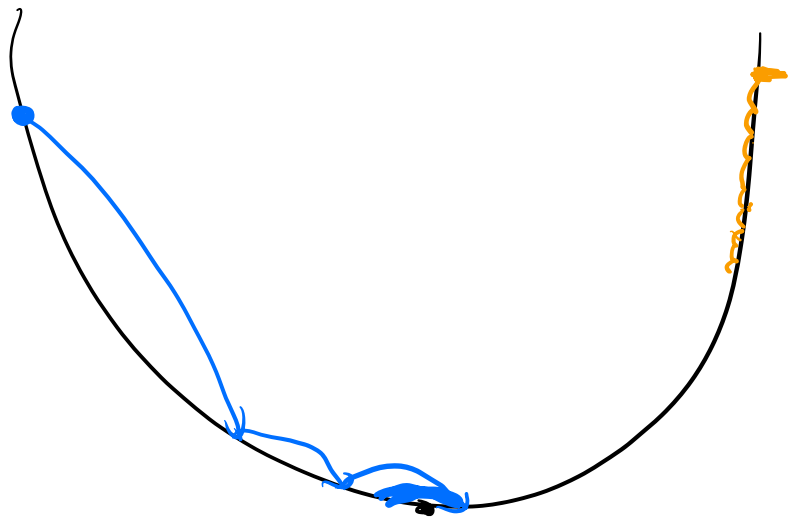
$w \leftarrow -\lambda \cdot dw$

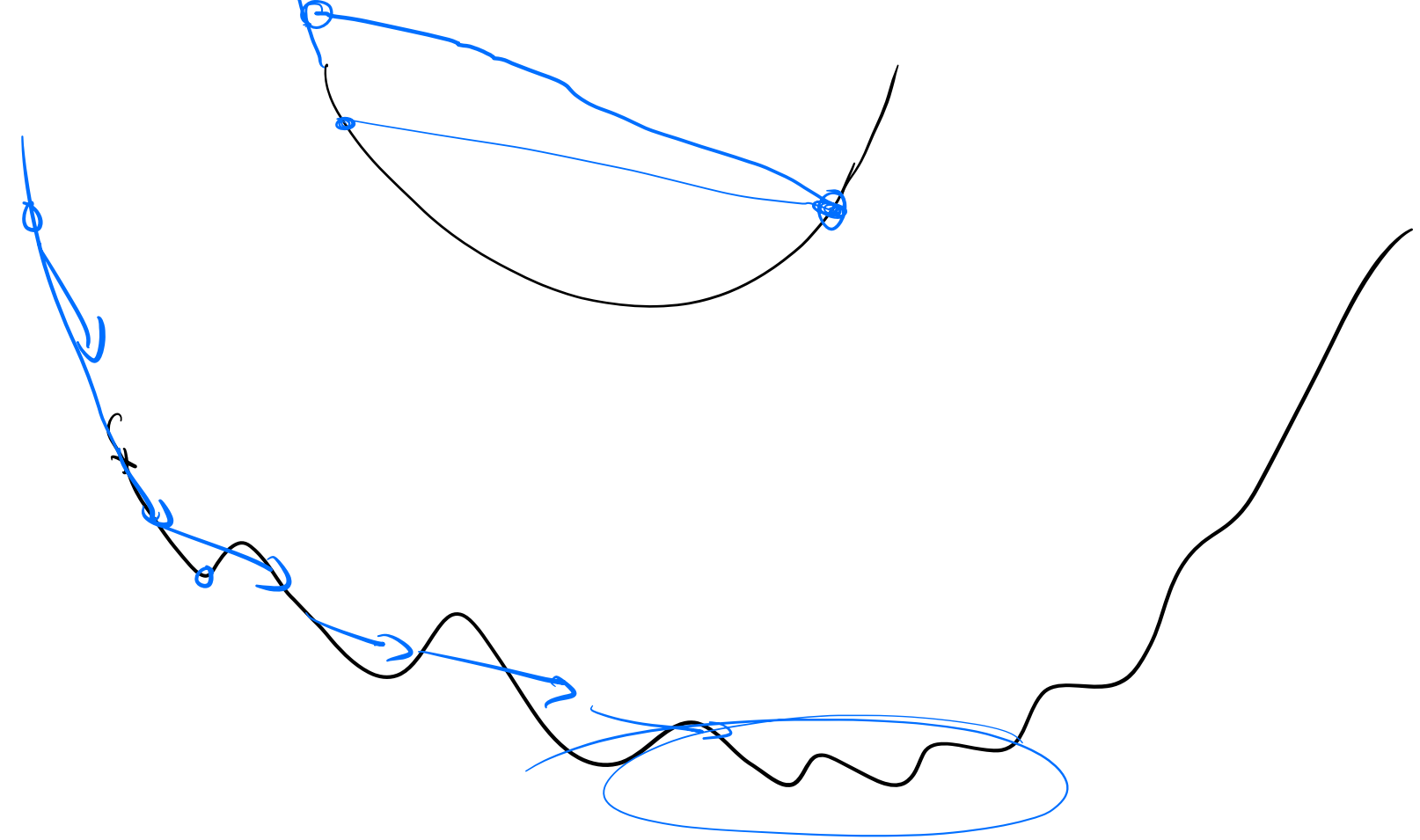
$b \leftarrow -\lambda \cdot db$

λ:

too big

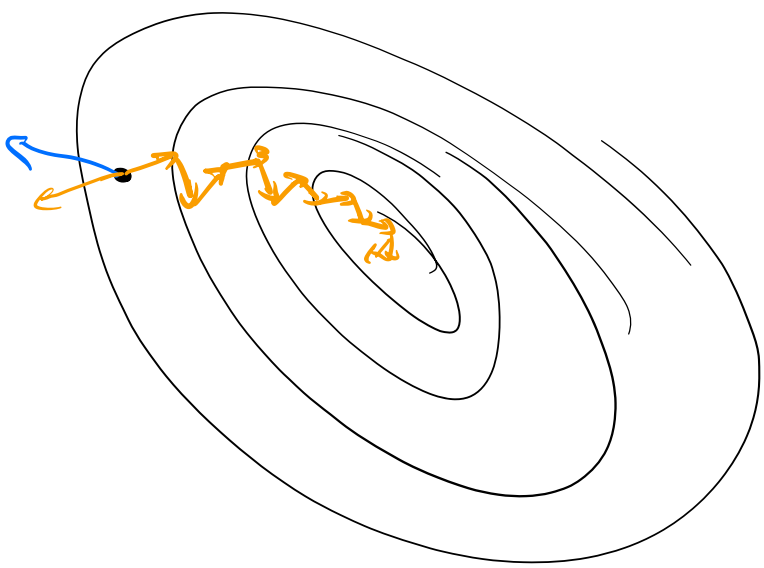
too small





Stochastic gradient descent

1 random pt per step



Minibatch GD

Choose batch of pts