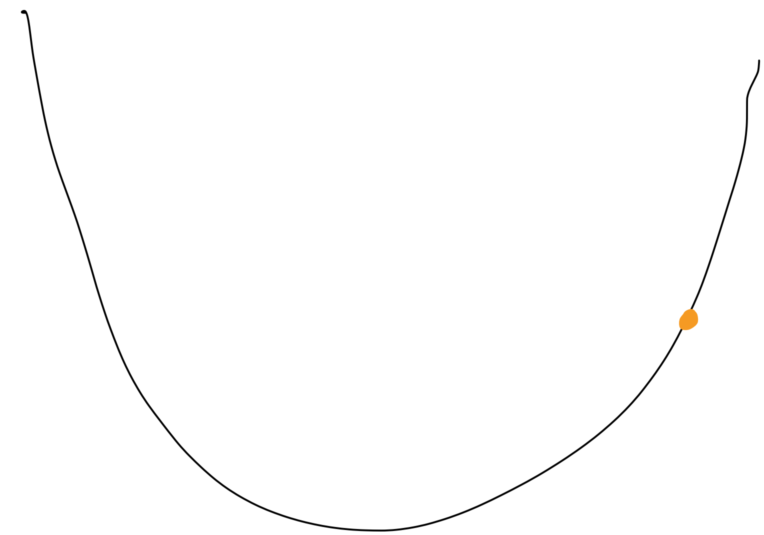
01/31/2023

$$f(\vec{v}) = v_1^2 + 2v_2 + 6v_1v_2$$

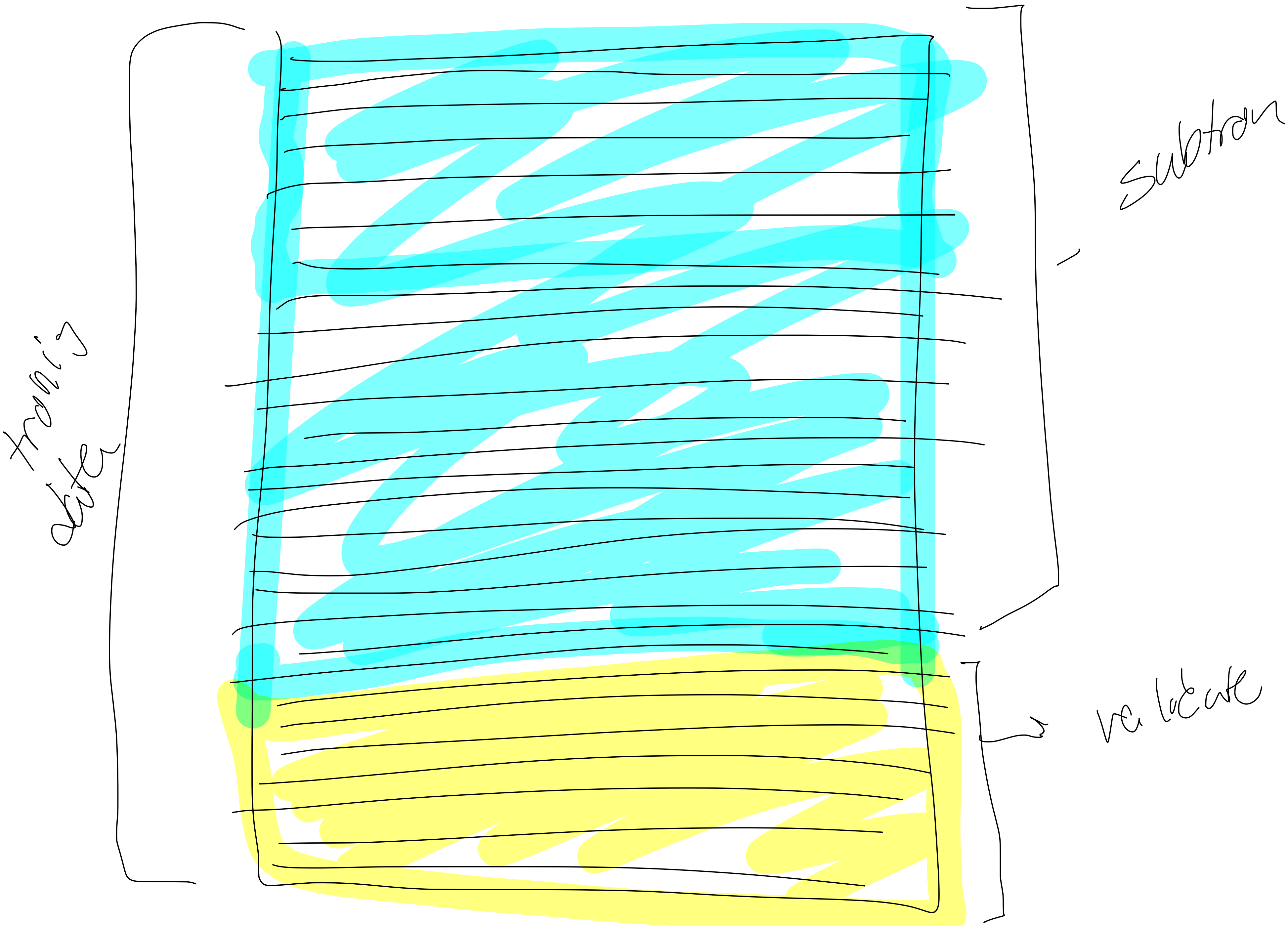
$$\nabla f = \begin{bmatrix} \partial f / \partial v_1 \\ \partial f / \partial v_2 \end{bmatrix}$$



norms

$$\|a\|_2 = \sqrt{\sum (a_i^2)} = \sqrt{\langle a, a \rangle}$$

$$\|a\|_1 = \sum |a_i|$$


training data
subtrain
validate

$$f(x) = 4x_1 + 3x_2 - 6x_1x_2 + 8x_3$$

$$\nabla f = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \partial f / \partial x_3 \end{bmatrix} = \begin{bmatrix} 4 - 6x_2 \\ 3 - 6x_1 \\ 8 \end{bmatrix}$$

$g(\vec{u})$ gradient descent

$$\vec{u}^{(t+1)} = \vec{u}^{(t)} - \alpha \nabla g(\vec{u}^{(t)})$$

$$\left\| \vec{u}^{(t+1)} - \vec{u}^{(t)} \right\|$$

$$\ell(\vec{a}, b, \lambda) = \left[\frac{1}{N}\sum_{i=1}^{N} \max(0, 1 - y_i(\vec{a}^T x_i + b))\right] + \lambda\frac{\|a\|^2}{2}$$

$$\ell(\vec{u}) = \left[\frac{1}{N}\sum_{i=1}^{N} \ell_i(\vec{u})\right] + \ell_0(\vec{u}) \qquad \vec{u} = \begin{bmatrix} b \\ \uparrow \\ a \\ \downarrow \end{bmatrix}$$

$$-\nabla\ell(\vec{u}) = -\left(\left[\underbrace{\frac{1}{N}\sum_{i=1}^{N}}_{\text{over all elements in dataset}} \nabla\ell_i(\vec{u})\right] + \nabla\ell_0(\vec{u})\right)$$

initialize $\vec{w}$ randomly

repeat until convergence:

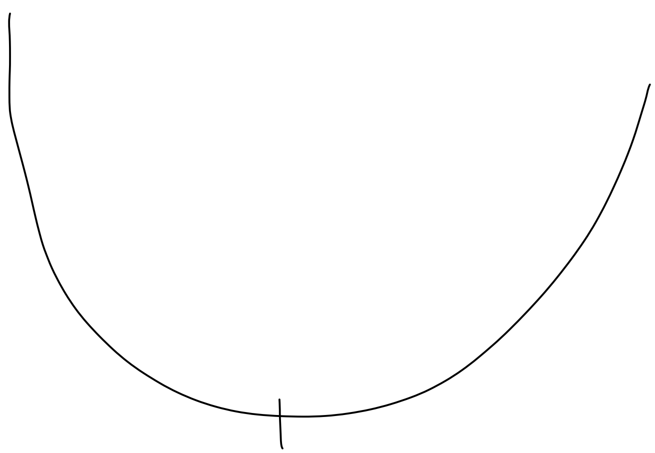
randomly shuffle samples

loop through samples and update weights:

$$\vec{u}^{(t+1)} = u^{(t)} - \alpha \nabla \ell(\vec{u}^{(t)})$$

alpha constant

gradient



Mini-batch gradient descent

choose fixed:

N_b = # of data points per batch

repeat until convergence

for N/N_b

grab random batch D of size N_b

update $\vec{u}$

$$\vec{u}^{(t+1)} = \vec{u}^{(t)} - \alpha\left[\frac{1}{N_b}\sum_{i\in D}\nabla l_i(\vec{u}^{(t)})) + \nabla \lambda l_0(\vec{u}^{(t)})\right]$$