APRML : Assignment 1

Anshuman Suri (2014021)

2. (A) *L1, L2, L_elastic norm:*

• $Z = \sigma(W_1 x + b_1)$, $\hat{X} = W_2 Z + b_2$

Standard error + regularizer, $e = \frac{1}{2}\|X - \hat{X}\|_2^2 + \lambda_1(|W_1| + |W_2|) + \lambda_2(\|W_1\|^2 + \|W_2\|^2)$

Assuming regularization for both sets of weights

L_1 & L_2 can be derived using elastic net (setting λ_1 or λ_2 as 0)

For elastic:

$$\frac{\partial e}{\partial W_2} = \underbrace{\frac{2.1}{2}(\hat{X} - X).(Z^T)}_{\text{normal part}} + \underbrace{\lambda_1 \, sign(W_2) + 2\lambda_2 W_2}_{\text{reg part}}$$

$$\frac{\partial e}{\partial W_1} = \underbrace{\frac{2.1}{2}(\hat{X} - X) W_2 . Z(1-Z).X}_{\text{normal part}} + \underbrace{\lambda_1 \, sign(W_1) + 2\lambda_2 W_1}_{\text{reg. part}}$$

Since regularizer is not on biases, $\frac{\partial e}{\partial b_1}$ & $\frac{\partial e}{\partial b_2}$ are unaffected.

These rules can therefore be used for back-prop.

For L_1 norm, set $\lambda_1 = 1$ & $\lambda_2 = 0$ in above

For L_2 norm, set $\lambda_1 = 0$ & $\lambda_2 = 1$ in above

L_trace norm:

$z = \sigma(w_1 x + b_1)$, $\hat{x} = w_2 z + b_2$

$e = \frac{1}{2} \|x - \hat{x}\|^2 +$ sum(eigenvalues (w_1)) + sum (eigenvalues (w_2))

sum(eigenvalues(X)) = $tr(\Sigma)$, where
$X = U \Sigma V^T$

$X = U \Sigma V^T$

$XV = U \Sigma V^T V = U \Sigma$

$U^T X V = U^T U \Sigma = \Sigma$

$\therefore tr(\Sigma) = tr(U^T X V)$

According to matrix cookbook, $\frac{\partial\, tr(AXB)}{\partial X} = (BA)^T$

Here we need $\frac{\partial\, tr(U^T X V)}{\partial X}$

However, the eigenvalues = 0 's eigenvectors in A & B will not affect the derivative. So, we consider only the part of U & V with $\lambda > 0$ i.e.,

$X = U' \Sigma' V'^T$ → cropped accordingly
↳ eigenvalues > 0

↓
cropped accordingly $\therefore \frac{\partial\, tr(U'^T \cancel{\Sigma} X V')}{\partial X} = \frac{\partial\, tr(U'^T X V')}{\partial x}$

$\{$as $tr(\Sigma) = tr(\Sigma')\}$

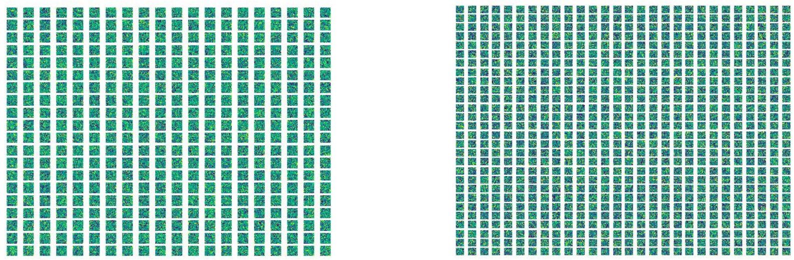
$\therefore \frac{\partial\, tr(U'^T X V')}{\partial X} = (V' U'^T)^T = U' V'^T$

Using this, $\frac{\partial e}{\partial w_2} = \frac{2 \cdot 1}{2}(\hat{x} - x) z^T + U_2' V_2'^T$,

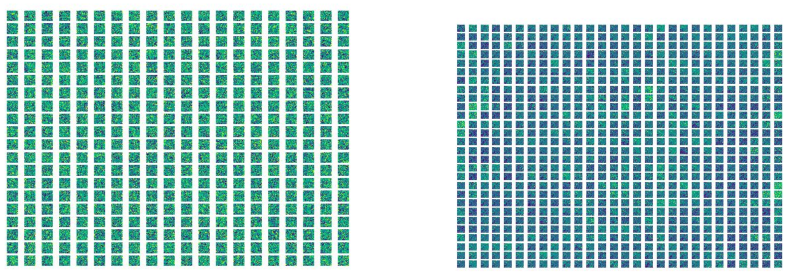
$\frac{\partial e}{\partial w_1} = \frac{2 \cdot 1}{2} w_2\, z(1-z) x + U_1' V_1'^T$

(B) *Visualized weights*

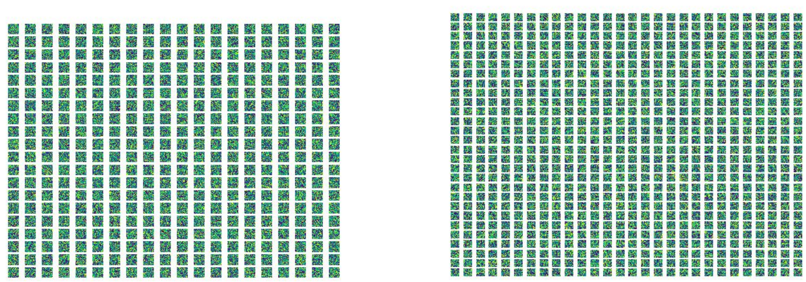
L1 norm (83.17% testing accuracy)



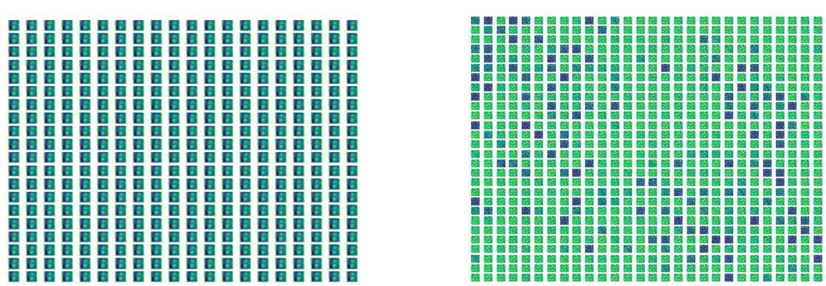
L2 norm (84.13% testing accuracy)



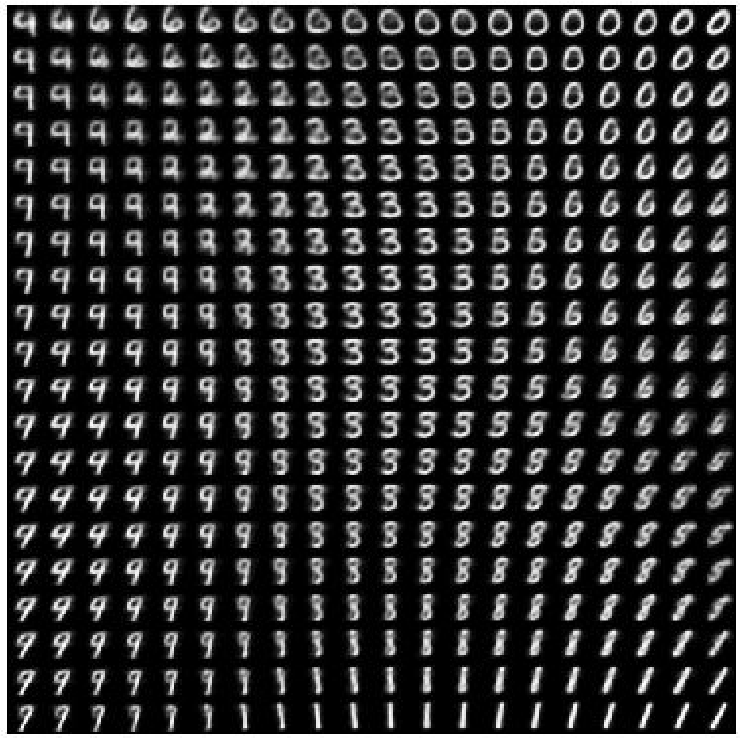
L-elastic (0.5, 0.5) norm (78.8% testing accuracy)



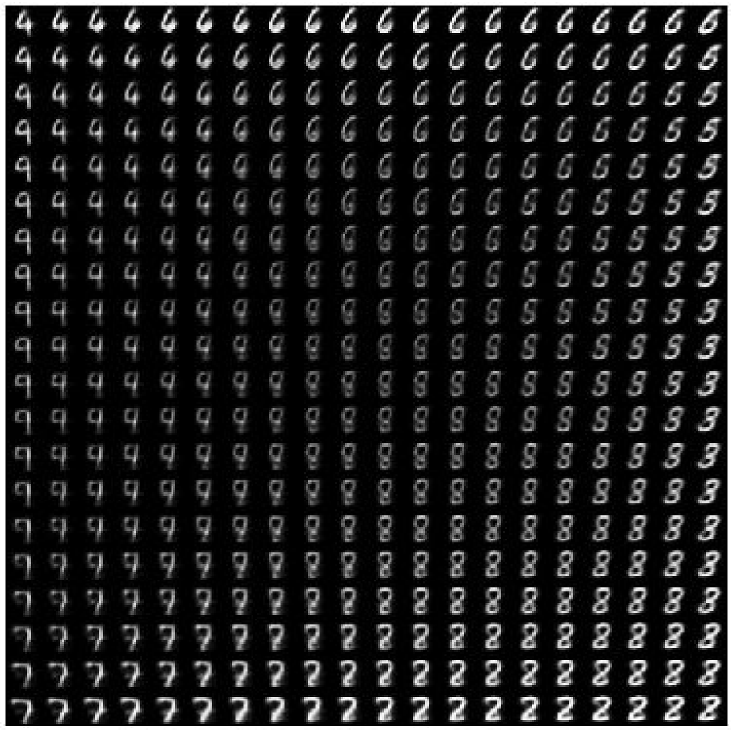
L-trace norm (46.47% testing accuracy)



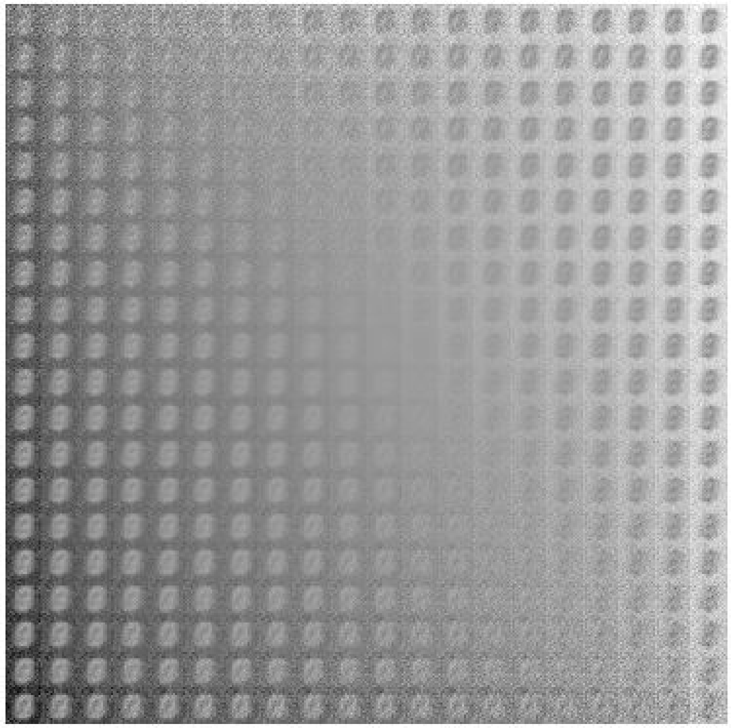
3. Bernoulli, 2D (Final loss 150.44)



Bernoulli, 5D (Final loss 119.06)



Gaussian, 2D (Final loss 45.27)



Gaussian, 5D (Final loss:)