Q1: (a) You can find graphs in attached directory.

Note: VAE1 is trained with (X-constructed - X)^2 as re construction Loss

VAE2 is trained with Cross Entropy Loss

VAE3 is trained with BCELoss

(b) You can find graphs in attached directory

(C): After training, both VAE and GAN have a better performance, the generator Error keeps decreasing. How ever, in GAN, discriminator Loss increases, which is expected. Discriminator Error increasing implies that it's harder to distinguish fake data and real data.

Contrast: Although both methods' performance improves while learning, the pattern they Learn varies.

VAE: VAE Compressed data to Lower dimension: So between two epochs, a slight change in Lower dimension could make images significantly different.

GAN: Incontrast, between two epoch, GAN's result looks much similar.

Question 2: First, we focus on the inner part. $\max_{\forall j \|z_j\|_2 \le \lambda} \|(x+z)w-y\|_2$

$$\max_{\forall j, \|z_j\|_2 \le \lambda} \|(x+z)w-y\|_2 = \max_{\forall j, \|z_j\|_2 \le \lambda} \|Xw-y+zw\|_2$$

$$= \max_{\forall j, \|z_j\|_2 \le \lambda} \max_{\|u\| \le 1} (Xw-y+Zw)^T u$$

$n \times d$ $d \times 1$

$$= \max_{\forall j, \|z_j\|_2 \le \lambda} \max_{\|u\| \le 1} (X \cdot w)^T u - y^T u + (zw)^T \cdot u$$

$$= \max_{\|u\| \le 1} \max_{\forall j, \|z_j\|_2 \le \lambda} (X \cdot w)^T u - y^T u + (z \cdot w)^T u \text{ (swap two max conditions)}$$

$$= \max_{\|u\| \le 1} (X \cdot w)^T u - y^T u + \max_{\forall j, \|z_j\|_2 \le \lambda} (z \cdot w)^T u \text{ (} z_j \text{ has nothing to do with } X, y \text{)}$$

$$= \max_{\|u\| \le 1} X \cdot w \cdot u - yu + \max_{\forall j, \|z_j\|_2 \le \lambda} w^T \cdot z^T u$$

$$= \max_{\|u\| \le 1} X \cdot w \cdot u - yu + \max_{\forall j, \|z_j\|_2 \le \lambda} \left(\sum_{i=1}^{d} w_i^T z_i^T\right) u \text{ (by Matrix multiplication)}$$

$$= \max_{\|u\| \le 1} x \cdot w \cdot u - y \cdot u + \max_{\forall j, \|A_j\|_2 \le 1} \left(\lambda \sum_{i=1}^{d} w_i^T A_i^T\right) \cdot u \quad \left(\text{Let } A_i = \frac{z_i}{\lambda}\right)$$

$$= \max_{\|u\| \le 1} x \cdot w \cdot u - y \cdot u + \max_{\forall j, \|A_j\| \|u\| \le 1} \left(\lambda \sum_{i=1}^{d} w_i^T \cdot A_i^T u\right) \quad (w_i^T \text{ is a scalar, } A_i^T \cdot u \le \|A_i\| \|u\| \le 1)$$

$$= \max_{\|u\| \le 1} x \cdot w \cdot u - y \cdot u + \lambda \sum_{i=1}^{d} w_i \quad \text{(By using Cauthy inequality reversely)}$$

$$= \|Xw - y\|_2 + \lambda \|w\|_1 \quad \text{(By using Cauthy inequality reversely)}$$