1. a) $E[\bar{x}] \in \mathbb{R}^d$

b) $E[\bar{x}] = E[\frac{1}{n}\sum_{i=1}^{n} x_i] = \frac{1}{n}(\sum_{i=1}^{n} E[x_i]) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}\cdot n\mu \boxed{= \mu}$

c) $Var[\bar{x}] = Var[\frac{1}{n}\sum_{i=1}^{n} x_i] = (\frac{1}{n})^2 Var[\sum_{i=1}^{n} x_i] = \frac{1}{n^2}\sum_{i=1}^{n} Var[x_i] = \frac{1}{n^2}\sum_{i=1}^{n}\Sigma = \frac{1}{n^2}\cdot n\Sigma = \boxed{\frac{\Sigma}{n}}$

d) $E[\bar{y}] = E[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d} x_{ij}] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d} E[x_{ij}] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d}\mu_i \boxed{= \frac{1}{n}\sum_{i=1}^{n} d\mu_i}$

e) $Var[\bar{y}] = Var[\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{d} x_{ij}] = (\frac{1}{n})^2 Var[\sum_{i=1}^{n}\sum_{j=1}^{d} x_{ij}] = \frac{1}{n^2}[\sum_{i=1}^{n} Var(\sum_{j=1}^{d} x_{ij})]$

$= \frac{1}{n}\sum_{j=1}^{d} Var[x_j] + \sum_{j=1}^{d}\sum_{k=1}^{d} Cov(x_j, x_k)_{j\neq k} = \frac{1}{n}\sum_{j=1}^{d}\sum_{k=1}^{d}\Sigma_{jk}$

2. a) No, since we are missing a rank our l_2 norm is not convex

b) $X = U\Sigma V^T$ $\quad U = \begin{bmatrix} \sqrt{2}/2 & 0 & -\sqrt{2}/2 \\ 0 & 1 & 0 \\ \sqrt{2}/2 & 0 & \sqrt{2}/2 \end{bmatrix}$ $\quad V^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $\quad \Sigma = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

$Xw = y$

$U\Sigma V^T w = y$ $\qquad$ Σ not invertible so Σ^+ pseudoinverse.

$w = V\Sigma^+ U^T y$

$$\therefore \; w = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & 0 & \sqrt{2}/2 \\ 0 & 1 & 0 \\ -\sqrt{2}/2 & 0 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1/\sqrt{2} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 1 \\ 0 \end{bmatrix} \boxed{= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}}$$

c) With $w_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$, we get $\boxed{w_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}$

With $w_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, we get $\boxed{w_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}$

The better solution is when $w_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ as then $Xw = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = y$.

This is a different solution than SVD, but still gives us a minimum l_2^2

d) As we increase w_0, further we get from an accurate w in 2 iterations.

e) $\|Xw - y\|_2^2 + \lambda \|w\|_2^2 = (Xw - y)^T (Xw - y) + \lambda w^T w$

$= w^T X^T X w - w^T X^T y + y^T X w + y^T y + \lambda w^T w$

$\therefore \nabla_w \|Xw - y\|_2^2 + \lambda \|w\|_2^2 = \nabla_w w^T X^T X w - \nabla_w w^T X^T y + \nabla_w y^T X w + \nabla_w y^T y + \nabla_w \lambda w^T w$

$= 2X^T X w - 2X^T y + 2\lambda w \Rightarrow 0$

$2X^T X w + 2\lambda w = 2X^T y$

$(X^T X + \lambda I) w = X^T y$

$w = (X^T X + \lambda I)^{-1} X^T y$

At $\lambda = 10^{-16}$: $w = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$

$\lambda = 10^{-2}$: $w = \begin{bmatrix} 0.99502488 \\ 0.99009901 \\ 0 \end{bmatrix}$

$\lambda = 10^{0}$: $w = \begin{bmatrix} 0.6\overline{6} \\ 0.5 \\ 0 \end{bmatrix}$

We see that as $\lambda \to 0$, we get closer to the same w as our l_2^2 norm