1. a) $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 3x_2 \\ x_2 \\ x_3 \\ 2x_5 \\ x_5 \end{pmatrix}$ Let $x_2 = a$, $x_3 = b$, $x_5 = c$ then $v = a\begin{pmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + c\begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 1 \end{pmatrix}$

the subspace is spanned by $v_1 = \begin{pmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 1 \end{pmatrix}$

We'll use the gram-schmidt process on these vectors to orthonormalize them

Normalize v_1: $u_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{3^2+1^2}}\begin{pmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$

Next, we subtract v_2's projection onto u_1 from v_2 (orthogonalize)

$u_2' = v_2 - \text{Proj}_{u_1}(v_2)$, $\text{Proj}_{u_1}(v_2) = \frac{v_2 \cdot u_1}{\underbrace{u_1 \cdot u_1}_{\text{this}=1}} u_1 = (v_2 \cdot u_1)u_1 = \left(\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \cdot \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}\right) u_1 = (0)u_1 = 0$, $u_2' = v_2$

v_2 is already normalized, so $u_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$

Normalize v_3 $\rightarrow \frac{v_3}{\|v_3\|} = \frac{1}{\sqrt{5}}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 1 \end{pmatrix}$ We can see that v_3 is already orthogonal to v_1 and v_2, so no need to orthogonalize

Our orthonormal basis is: $\boxed{u_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, u_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, u_3 = \frac{1}{\sqrt{5}}\begin{pmatrix} 0 \\ 0 \\ 0 \\ 2 \\ 1 \end{pmatrix}}$

b) The following equation must only hold when $c_1, c_2, c_3 = 0$

$c_1u_1 + c_2u_2 + c_3(u_2 + u_3) = 0$

$\rightarrow c_1u_1 + c_2u_2 + c_3u_2 + c_3u_3$

$c_1u_1 + (c_2+c_3)u_2 + c_3u_3$

We know that u_1, u_2, u_3 are orthonormal to eachother, and multiplying their values by a constant still yields the same result $\therefore$ the only solution is $c_1, c_2, c_3 = 0$

2. a) $B = AA^T = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1\cdot1 + 2\cdot2 & 1\cdot2 + 2\cdot3 & 1\cdot3 + 2\cdot1 \\ 2\cdot1 + 3\cdot2 & 2\cdot2 + 3\cdot3 & 2\cdot3 + 3\cdot1 \\ 3\cdot1 + 1\cdot2 & 3\cdot2 + 1\cdot3 & 3\cdot3 + 1\cdot1 \end{pmatrix} = \begin{pmatrix} 5 & 8 & 5 \\ 8 & 13 & 9 \\ 5 & 9 & 10 \end{pmatrix}$

b) The singular values of A are the square roots of the eigenvalues of AA^T.

$B = AA^T$, so the singular values of A are the square roots of the eigenvalues of B.

Since B has $\lambda_1 = 25, \lambda_2 = 3, \lambda_3 = 0$

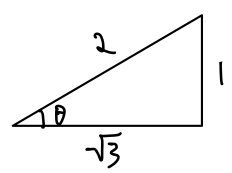
A has $\sigma_1 = 5, \sigma_2 = \sqrt{3}, \sigma_3 = 0$

Σ will match the dimensions of A (3×2)

So $\boxed{\Sigma = \begin{pmatrix} 5 & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{pmatrix}}$ Σ is diagonal with singular values

c) compare U to R_θ

we can see $\cos\theta = \sqrt{3}/2$, $\sin\theta = 1/2$ ∴ the matrices match



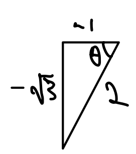
$\theta = \cos^{-1}(\sqrt{3}/2) = \pi/6$

$\begin{pmatrix} \cos(\frac{\pi}{6}) & -\sin(\pi/6) \\ \sin(\pi/6) & \cos(\pi/6) \end{pmatrix} = \begin{pmatrix} 3/\sqrt{2} & -1/2 \\ 1/2 & 3/\sqrt{2} \end{pmatrix} = U, \theta = \frac{\pi}{6}$

---

compare V_T to R_θ

we can see $\cos\theta = -\frac{1}{2}$, $\sin\theta = -\frac{\sqrt{3}}{2}$ ∴ the matrices match



$\theta = \cos^{-1}(-1/2) = 4\pi/3$

$\begin{pmatrix} \cos(4\pi/3) & -\sin(4\pi/3) \\ \sin(4\pi/3) & \cos(4\pi/3) \end{pmatrix} = \begin{pmatrix} -1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & -1/2 \end{pmatrix} = V, \theta = \frac{4\pi}{3}$

3.a) $\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3} \right)$

$$\frac{\partial f}{\partial x_1} = \frac{\partial}{\partial x_1}(5x_1^2 + 3x_2^2 + 2x_3^2 + 4x_1x_2 - 2x_1x_3 + 6x_2x_3)$$
$$= 10x_1 + 4x_2 - 2x_3$$

$$\frac{\partial f}{\partial x_2} = 6x_2 + 4x_1 + 6x_3, \quad \frac{\partial f}{\partial x_3} = 4x_3 - 2x_1 + 6x_2$$

$$\nabla f = (10x_1 + 4x_2 - 2x_3, 6x_2 + 4x_1 + 6x_3, 4x_3 - 2x_1 + 6x_2)$$

$$H(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_1 \partial x_3} \\ \frac{\partial^2 f}{\partial x_2 x_1} & \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x_2 \partial x_3} \\ \frac{\partial^2 f}{\partial x_3 x_1} & \frac{\partial^2 f}{\partial x_3 \partial x_2} & \frac{\partial^2 f}{\partial x_3^2} \end{pmatrix}$$

$$\frac{\partial^2 f}{\partial x_1^2} = 10, \quad \frac{\partial^2 f}{\partial x_1 x_2} = 4, \quad \frac{\partial^2 f}{\partial x_1 x_3} = -2$$
$$\frac{\partial^2 f}{\partial x_2 x_1} = 4, \quad \frac{\partial^2 f}{\partial x_2^2} = 6, \quad \frac{\partial^2 f}{\partial x_2 \partial x_3} = 6 \qquad H = \begin{pmatrix} 10 & 4 & -2 \\ 4 & 6 & 6 \\ -2 & 6 & 4 \end{pmatrix}$$
$$\frac{\partial^2 f}{\partial x_3 x_1} = -2, \quad \frac{\partial^2 f}{\partial x_3 \partial x_2} = 6, \quad \frac{\partial^2 f}{\partial x_3^2} = 4$$

b) $g(x) = f(\vec{x}_0) + \frac{1}{1!}(\vec{x} - \vec{x}_0)^T \frac{\partial f}{\partial \vec{x}}(\vec{x}_0) + \frac{1}{2!}(\vec{x} - \vec{x}_0)^T \frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^T}(\vec{x} - \vec{x}_0)$

$$= f(\vec{x}_0) + \nabla f(\vec{x}_0)(\vec{x} - \vec{x}_0)^T + \frac{1}{2}(\vec{x} - \vec{x}_0)^T Hf(\vec{x}_0)(\vec{x} - \vec{x}_0)$$

$$f(\vec{x}_0) = 0, \quad \nabla f(\vec{x}_0)(\vec{x} - \vec{x}_0)^T = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}(\vec{x} - \vec{x}_0)^T = 0$$

$$\frac{1}{2!}(\vec{x} - \vec{x}_0)^T \frac{\partial^2 f}{\partial \vec{x} \partial \vec{x}^T}(\vec{x} - \vec{x}_0) = \frac{1}{2}(x_1, x_2, x_3) \begin{pmatrix} 10 & 4 & -2 \\ 4 & 6 & 6 \\ -2 & 6 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$
$$= \frac{1}{2}(x_1, x_2, x_3) \begin{pmatrix} 10x_1 + 4x_2 - 2x_3 \\ 4x_1 + 6x_2 + 6x_3 \\ -2x_1 + 6x_2 + 4x_3 \end{pmatrix}$$

$$= \frac{1}{2}(10x_1^2 + 4x_1x_2 - 2x_1x_3 + 4x_1x_2 + 6x_2^2 + 6x_2x_3 - 2x_1x_3 + 6x_2x_3 + 4x_3^2$$

$$\boxed{g(x) = \frac{1}{2}(10x_1^2 + 6x_2^2 + 4x_3^2 + 8x_1x_2 - 4x_1x_3 + 12x_2x_3)}$$

c) We can use sylvester's criterion. Check determinants of the upper left 1x1, 2x2, and 3x3 submatrices. PSD -> all values are +
NSD -> " " −
Neither -> + or −

$$H = \begin{pmatrix} 10 & 4 & -2 \\ 4 & 6 & 6 \\ -2 & 6 & 4 \end{pmatrix} \qquad 1\text{x}1 \text{ case: } |10| = 10$$

2x2 case: $\begin{vmatrix} 10 & 4 \\ 4 & 6 \end{vmatrix} = 10 \cdot 6 - 4 \cdot 4 = 60 - 16 = 44$

3x3 case: $\begin{vmatrix} 10 & 4 & -2 \\ 4 & 6 & 6 \\ -2 & 6 & 4 \end{vmatrix} = 10 \begin{vmatrix} 6 & 6 \\ 6 & 4 \end{vmatrix} - 4 \begin{vmatrix} 4 & 6 \\ -2 & 4 \end{vmatrix} + (-2) \begin{vmatrix} 4 & 6 \\ -2 & 6 \end{vmatrix}$

$= 10(24-36) - 4(16+12) - 2(24+12)$

$= -120 - 112 - 72 = -304$

Our values are: 10, 44, -304. Since we have positive & negative, f is neither

---

4. we can use the first criteria

$ReLU(tx+(1-t)y) \leq tReLU(x) + (1-t)ReLU(y)$

Case 1: $x \geq 0$, $y \geq 0$

$ReLU(x) = x$, $ReLU(y) = y$

$ReLU(tx+(1-t)y) \leq tReLU(x) + (1-t)ReLU(y)$

$tx + (1-t)y \leq tx + (1-t)y$ ✓

Case 2: $x \leq 0$, $y \leq 0$

$Relu(x) = 0$, $Relu(y) = 0$

$Relu(0) \leq t \cdot 0 + (1-t) \cdot 0$

$0 \leq 0$ ✓

Case 3: $x \geq 0$, $y < 0$

$ReLU(x) = x$, $ReLU(y) = 0$

$ReLU(tx+(1-t)y) \leq tx$

we know $(1-t)y$ is always ≤ 0, so this reduces the value of tx in the LHS by z $(z \geq 0)$

$ReLU(tx - z) \leq tx$

$tx - z \leq tx$ ✓

∴ the ReLU function is convex

5. a)
b)

```
Degree 1
Linear Regression average MSE: 31.605042346619875
Ridge Regression average MSE: 31.551288974302498
Ridge Regression best alpha: 10

Degree 2
Linear Regression average MSE: 23.474324700591257
Ridge Regression average MSE: 22.83762009481931
Ridge Regression best alpha: 100

Degree 3
Linear Regression average MSE: 27.97201681936091
Ridge Regression average MSE: 22.99773310403309
Ridge Regression best alpha: 100

Degree 4
Linear Regression average MSE: 30.10917225220887
Ridge Regression average MSE: 22.578216443022605
Ridge Regression best alpha: 100

Degree 5
Linear Regression average MSE: 40.60410247000668
Ridge Regression average MSE: 24.809472378601008
Ridge Regression best alpha: 10

Overall best parameters
Linear Regression best degree: 2
Ridge Regression best degree: 4
Ridge Regression best alpha: 100
```

c)

```
Linear Regression error: 36.48367311568473
Ridge Regression error: 19.482105431320235
```

D)

```
"""
Questions

Part C
Ridge regression has a better MSE score, meaning in this case, it generalizes to new data better. In general, Ridge
regression is less prone to overfitting due to penalization of high degrees, this means that for general datasets,
ridge regression will usually generalize to new data better.

Part D
High degree polynomials can overfit due to their flexibilty. Ridge penalizes this by adding a regularization term
(alpha) which is proportional to the sum of the squared coefficients. The larger the alpha & degree, the stronger
the penalization. The choice of alpha can change how closely the model fits the data. A large degree can lead to
overfitting, and a large alpha can lead to underfitting. The overfitting can be countered by a large enough alpha.
"""
```