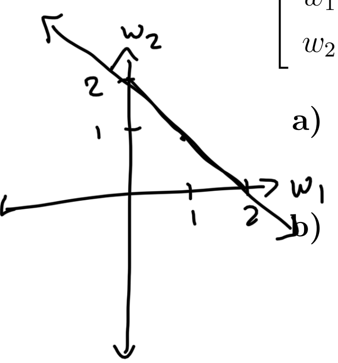
b) The solution is unique. $||w||_2^2$ gives a circle in the w_1-w_2 plane and that circle will only intersect the set of all w once, so there is only one solution that satisfies $\min_w ||w||_2^2$ and $Xw = y$.

CS/ECE/ME532 Activity 9

Estimated Time: 25 minutes for P1, 30 minutes for P2

1. Consider the system of linear equations $\boldsymbol{Xw} = \boldsymbol{y}$ where $\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}$, $\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ and $\boldsymbol{y} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$.

$\begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} w_1+w_2 \\ -2w_1+-2w_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$



a) Sketch the set of all $\boldsymbol{w}$ that satisfy $\boldsymbol{Xw} = \boldsymbol{y}$ in the w_1-w_2 plane. Is the solution unique? What is the value of the squared error $\min_{\boldsymbol{w}} ||\boldsymbol{Xw} - \boldsymbol{y}||_2^2$? $= w = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

The solution is not unique.

b) Use your sketch to find the $\boldsymbol{w}$ of minimum norm that satisfies the system of equations: $\min_{\boldsymbol{w}} ||\boldsymbol{w}||_2^2$ subject to $\boldsymbol{Xw} = \boldsymbol{y}$. Is this solution unique? What makes it unique? What is the value of the squared error $||\boldsymbol{Xw} - \boldsymbol{y}||_2^2$ at this solution? What is the value of $||\boldsymbol{w}||^2$? *Hint:* The equation $||\boldsymbol{w}||_2^2 = c$ describes a circle in $\mathbb{R}^2$ with radius $\sqrt{c}$.

$w = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $||w||_2^2 = (\sqrt{1^2+1^2})^2 = 2$ $||Xw-y||_2^2 = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 \\ -4 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix} - \begin{bmatrix} 2 \\ -4 \end{bmatrix} = 0$

c) Algebraically find the $\hat{\boldsymbol{w}}$ that solves the Tikhonov-regularized (or ridge regression) problem $\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \{||\boldsymbol{Xw} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{w}||_2^2\}$ as a function of λ. *Hint:* Recall that $w = (X^TX + \lambda I)^{-1} X^T y$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

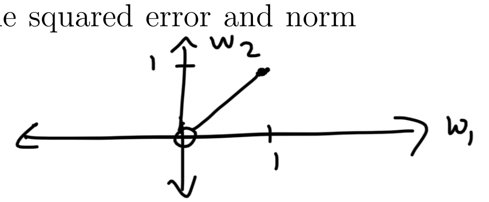
$\rightarrow \frac{1}{10\lambda+\lambda^2}\begin{bmatrix} 5+\lambda & -5 \\ -5 & 5+\lambda \end{bmatrix}$

$X^TX = \begin{bmatrix} 1 & -2 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix} = \begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix}$

$X^TX + \lambda I = \begin{bmatrix} 5+\lambda & 5 \\ 5 & 5+\lambda \end{bmatrix}$

$\frac{1}{(5+\lambda)^2 - 5^2} = \frac{1}{25+10\lambda+\lambda^2-25} = \frac{1}{10\lambda+\lambda^2}$

d) Sketch the set solution to the Tikhonov-regularized problem in the w_1-w_2 plane as a function of λ for $0 < \lambda < \infty$. (Consider the solution for different values of λ in that range.) Find the squared error $||\boldsymbol{Xw} - \boldsymbol{y}||_2^2$ and norm squared of the solution, $||\boldsymbol{w}||_2^2$ for $\lambda = 0$ and $\lambda = 5$. Compare the squared error and norm squared of the solution to those in part **b)**.



$w = \frac{1}{10\lambda+\lambda^2}\begin{bmatrix} 5+\lambda & -5 \\ -5 & 5+\lambda \end{bmatrix}\begin{bmatrix} 1 & -2 \\ 1 & -2 \end{bmatrix}\begin{bmatrix} 2 \\ -4 \end{bmatrix}$

$= \begin{bmatrix} \lambda & -2\lambda \\ \lambda & -2\lambda \end{bmatrix}\begin{bmatrix} 2 \\ -4 \end{bmatrix} = \begin{bmatrix} 10\lambda \\ 10\lambda \end{bmatrix} = \begin{bmatrix} \frac{10}{10+\lambda} \\ \frac{10}{10+\lambda} \end{bmatrix}$

$\lambda = 0 \rightarrow w = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \rightarrow ||w||_2^2 = 2$

$\rightarrow ||Xw-y||_2^2 = 0$

$w = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \end{bmatrix} \rightarrow ||w||_2^2 = \frac{8}{9}$

$\rightarrow ||Xw-y||_2^2 = \frac{20}{9}$

2. Let $\boldsymbol{X} = \begin{bmatrix} 1 & \gamma \\ 1 & -\gamma \\ 1 & -\gamma \\ 1 & \gamma \end{bmatrix}$.

a) Show that the columns of $\boldsymbol{X}$ are orthogonal to each other for any γ.

$x_1^T x_2 = 0$ $\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} x_2 = \gamma - \gamma - \gamma + \gamma = 0$

b) Express $\boldsymbol{X} = \boldsymbol{U\Sigma}$ where $\boldsymbol{U}$ is a 4-by-2 matrix with orthonormal columns and $\boldsymbol{\Sigma}$ is a 2-by-2 diagonal matrix (the non-diagonal entries are zero).

$X = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} 2 & 0 \\ 0 & 2\gamma \end{bmatrix}$

$X^T X = I$

if X has orthonormal columns

$w = (X^T X)^{-1} X^T y$

c) Express the solution to the least-squares problem $\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ as a function of $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{y}$.

$w = ((U\Sigma)^T (U\Sigma))^{-1} (U\Sigma)^T y$

d) Let $\boldsymbol{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. Find the weights $\boldsymbol{w}$ as a function of γ. What happens to $||\boldsymbol{w}||_2^2$ as $\gamma \to 0$?

$\begin{bmatrix} 1 & 1 & 1 & 1 \\ \gamma & -\gamma & -\gamma & \gamma \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2\gamma \end{bmatrix}$ As $\lambda \to 0$, $||w||_2^2$ gets closer to 2.

$\gamma = 0.1 \to \frac{2}{2(0.1)} = 10$

$\gamma = 10^{-8} = \frac{2}{2(10^{-8})} = 10^8$

e) The ratio of the largest to the smallest diagonal values in $\boldsymbol{\Sigma}$ is termed the condition number of $\boldsymbol{X}$. Find the condition number if $\gamma = 0.1$ and $\gamma = 10^{-8}$. Also find $||\boldsymbol{w}||_2^2$ for these two values of γ. $\gamma = 0.1$ $w = \begin{bmatrix} 2 \\ 0.2 \end{bmatrix}$ $||w||_2^2 = 4.4$ $\gamma = 10^{-8}$ $w = \begin{bmatrix} 2 \\ 2 \cdot 10^{-8} \end{bmatrix}$ $||w||_2^2 = 4 + (2 \cdot 10^{-8})^2$

f) A system of linear equations with a large condition number is said to be "ill-conditioned". One consequence of an ill-conditioned system of equations is solutions with large norms as you found in the previous part of this problem. A second consequence is that the solution is very sensitive to small errors in $\boldsymbol{y}$ such as may result from measurement error or numerical error. Suppose $\boldsymbol{y} = \begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix}$. Write $\boldsymbol{w} = \boldsymbol{w}_o + \boldsymbol{w}_\epsilon$ where $\boldsymbol{w}_o$ is the solution for arbitrary γ when $\epsilon = 0$ and $\boldsymbol{w}_\epsilon$ is the perturbation in that solution due to some error $\epsilon \neq 0$. How does the norm of the perturbation due to $\epsilon \neq 0$, $||\boldsymbol{w}_\epsilon||_2^2$, depend on the condition number? Find $||\boldsymbol{w}_\epsilon||_2^2$ for $\epsilon = 0.01$ and $\gamma = 0.1$ and $\gamma = 10^{-8}$. The smaller γ, the more apparent the difference due to adding ε will be.

$W_\varepsilon = \begin{bmatrix} 2+\varepsilon \\ 2r + \gamma\varepsilon \end{bmatrix}$

$W = \begin{bmatrix} 2 \\ 2r \end{bmatrix} + \begin{bmatrix} 2+\varepsilon \\ 2r + \gamma\varepsilon \end{bmatrix}$

$= \begin{bmatrix} 4+\varepsilon \\ 4\gamma + \gamma\varepsilon \end{bmatrix}$

g) Now apply ridge regression, i.e., Tikhonov reqularization. Solve for $\boldsymbol{w}_o$ and $\boldsymbol{w}_\epsilon$ as a function of λ. Find $||\boldsymbol{w}_o||_2^2$ and $||\boldsymbol{w}_\epsilon||_2^2$ for $\lambda = 0.1$, $\epsilon = 0.01$ and $\gamma = 0.1$ and $\gamma = 10^{-8}$. Comment on the impact of regularization.

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

$\gamma = 0.1$

$W = \begin{bmatrix} 4 + 0.01 \\ 0.4 + 0.001 \end{bmatrix}$

$= \begin{bmatrix} 4.01 \\ 0.401 \end{bmatrix}$ $||w||_2^2 = (4.01)^2 + (0.401)^2$

$\gamma = 10^{-8}$

$w = \begin{bmatrix} 4 + 0.01 \\ 4 \cdot 10^{-8} + 0.01 \cdot 10^{-8} \end{bmatrix}$

$||w||_2^2 = (4.01)^2 + (4 \cdot 10^{-8} + 0.01 \cdot 10^{-8})^2$