Machine Learning - 2020-2021
Universitat Pompeu Fabra

Problems 6B: Linear Regression and Regularisation

Goal

The goal of this practice is to understand the regularisation of the linear regression model and the Gradient Descent algorithm. Previously, we review and practice the differentiation of vectors and matrices.

Needed concepts

If $f(\mathbf{x})$ is a function of $x_1, \ldots, x_D$, then we define the derivative of $f(\mathbf{x})$ *with respect to the vector* $\mathbf{x}$ as the vector of derivatives: $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_D} \end{pmatrix}$. This is very useful for expressions as $y = \mathbf{x^T b} = (x_1, \ldots, x_D) \begin{pmatrix} b1 \\ \ldots \\ b_D \end{pmatrix} = x_1 b_1 + \ldots, b_D$. Instead of deriving with respect to $x_1, \ldots, x_D$ we organize the resuls in a vector form. The main rules are:

$\frac{\partial \mathbf{x}^T \mathbf{b}}{\partial \mathbf{x}} = \mathbf{b}$	$\frac{\partial \mathbf{b}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{b}$	$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{A}$	if $\mathbf{A}$ symmetric, $\frac{\partial \mathbf{x}^T \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}$ else: $\frac{\partial \mathbf{x}^T \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x}$

Exercises

Linear Regression and Regularisation

1. Consider the following training set $\left\{\left(x^{(i)}, y^{(i)}\right)\right\}_{i=1}^{5}$ with feature variables $\{x^{(1)} = 10, x^{(2)} = 10, x^{(3)} = 20, x^{(4)} = 35, x^{(5)} = 40\}$ and target variables $\left\{y^{(1)} = 10, y^{(2)} = 20, y^{(3)} = 20, y^{(4)} = 25, y^{(5)} = 35\right\}$.

   (a) *Closed-form solution*

   i. Estimate the parameters of a Linear Regression model using (1) the closed-form solution and (2) the closed form solution when a Ridge regression is applied (with $\lambda = 1$).

   ii. Compute the error in the training set for both models. Plot the training set and both models in the same figure.

   What differences can you see between the models? Which is the effect of the regularisation?

Solution:

The linear regression model is represented by a linear function $g(\mathbf{x}, \mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$ where $\mathbf{x} = (1, x_1, \cdots, x_D) \in \mathbb{R}^{D+1}$, with D the dimension of the training data. In this particular case, $g(x) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$, where $\mathbf{x} = (1, x)^{\mathrm{T}}$, thus $D = 1$.

To train a linear regression model, first, we should construct the design matrix:

$$\mathbf{X} = \begin{pmatrix} 1 & x^{(1)} \\ \vdots & \vdots \\ 1 & x^{(5)} \end{pmatrix} = \begin{pmatrix} 1 & 10 \\ 1 & 10 \\ 1 & 20 \\ 1 & 35 \\ 1 & 40 \end{pmatrix}$$

with target variable

$$\mathbf{y} = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(5)} \end{pmatrix} = \begin{pmatrix} 10 \\ 20 \\ 20 \\ 25 \\ 30 \end{pmatrix}.$$

Without regularisation

The closed-form solution is

$$\begin{aligned}\mathbf{w} &= \left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top\mathbf{y}\\ &= \left[\begin{pmatrix}1 & 1 & 1 & 1 & 1\\ 10 & 10 & 20 & 35 & 40\end{pmatrix}\begin{pmatrix}1 & 10\\ 1 & 10\\ 1 & 20\\ 1 & 35\\ 1 & 40\end{pmatrix}\right]^{-1}\begin{pmatrix}1 & 1 & 1 & 1 & 1\\ 10 & 10 & 20 & 35 & 40\end{pmatrix}\begin{pmatrix}10\\ 20\\ 20\\ 25\\ 30\end{pmatrix}\\ &= \begin{pmatrix}5 & 115\\ 115 & 3425\end{pmatrix}^{-1}\begin{pmatrix}105\\ 2775\end{pmatrix} = \frac{1}{780}\begin{pmatrix}685 & -23\\ -23 & 1\end{pmatrix}\begin{pmatrix}105\\ 2775\end{pmatrix} = \frac{3}{13}\begin{pmatrix}45\\ 2\end{pmatrix} \approx \begin{pmatrix}10.38\\ 0.46\end{pmatrix}\end{aligned}$$

Error:

$$\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{5}\left(g\left(x^{(i)}\right) - y^{(i)}\right)^2 = \frac{1}{2}\sum_{i=1}^{5}\left(w_0 + w_1x^{(i)} - y^{(i)}\right)^2 = \cdots \approx 26.92$$

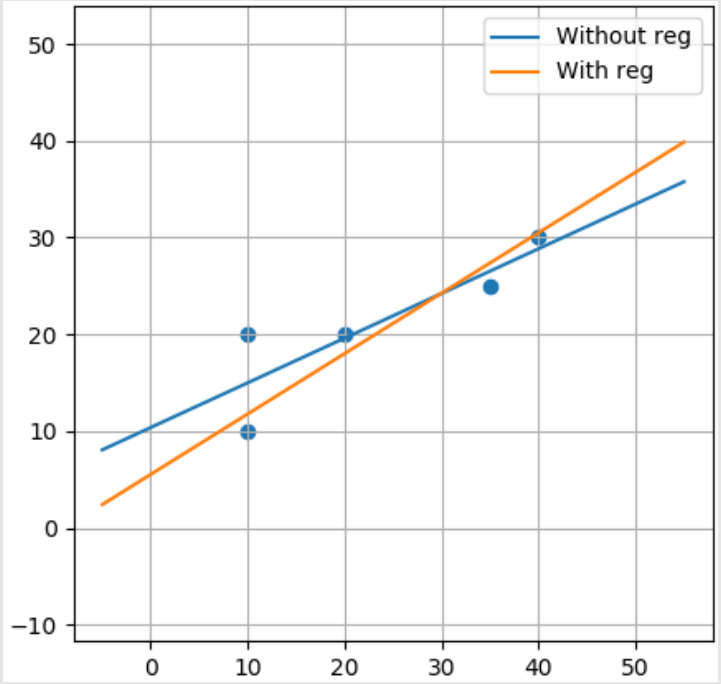
Without regularisation

The closed-form solution is

$$\mathbf{w} = \left(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{Id}\right)^{-1}\mathbf{X}^\top\mathbf{y} = \ldots\text{(same as above)}\cdots \approx \begin{pmatrix}5.54\\ 0.62\end{pmatrix}$$

Error:

$$\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{5}\left(g\left(x^{(n)}, \mathbf{w}\right) - y^{(n)}\right)^2 = \frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1x^{(n)} - y^{(n)}\right)^2 = \cdots \approx 40.29$$



(b) *Gradient descent by hand*: Compute a Gradient Descent update for a Linear regression model following the steps below.

i. Write the parametric form of a linear regression model $g(\mathbf{x}, \mathbf{w})$ and the error function that is minimised when fitting a linear model to the given data.

Solution:
Linear Regression: $g(\mathbf{x}, \mathbf{w}) = \mathbf{w}^{\mathrm{T}}\mathbf{x}$ where $\mathbf{x} = (1, x) \in \mathbb{R}^2$
Error:

$$\mathbb{E}(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{N}\left(g(\mathbf{x}^{(n)}, \mathbf{w}) - y^{(n)}\right)^2 = \frac{1}{2}\sum_{n=1}^{N}\left(\mathbf{w}^{\mathrm{T}}\mathbf{x}^{(n)} - y^{(n)}\right)^2 = \frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)^2$$

ii. Find the expression of the gradient descent update of the parameters $\mathbf{w}$. To do so, derive the error function from (i) with respect to each of the parameters w_j.

Solution:

$$\begin{aligned}\frac{\partial \mathbb{E}(\mathbf{w})}{\partial w_j} &= \frac{\partial}{\partial w_j}\frac{1}{2}\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)^2 \\ &= \sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)\frac{\partial}{\partial w_j}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)\end{aligned}$$

UPDATES:

$$\begin{aligned}w_0 &= w_0 - \alpha\frac{\partial \mathbb{E}(\mathbf{w})}{\partial w_0} \\ &= w_0 - \alpha\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)\frac{\partial}{\partial w_j}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right) \\ &= w_0 - \alpha\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)\end{aligned}$$

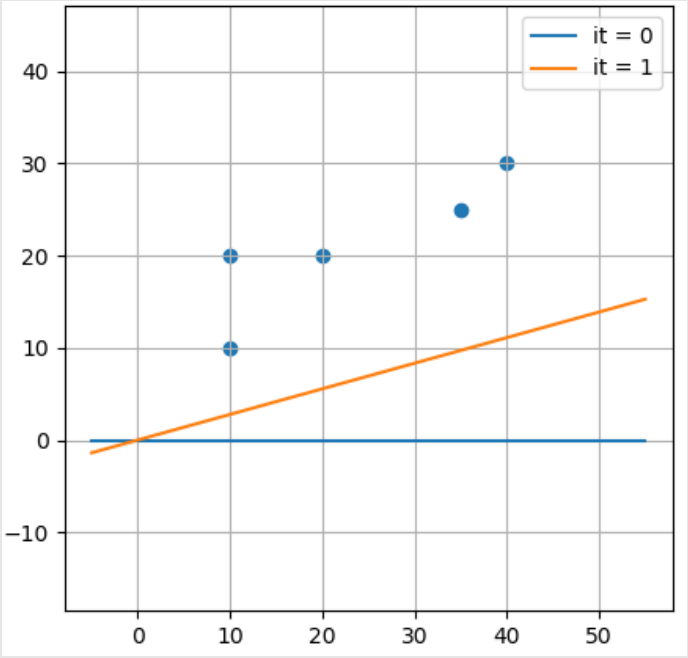
$$\begin{aligned}w_1 &= w_1 - \alpha\frac{\partial \mathbb{E}(\mathbf{w})}{\partial w_1} \\ &= w_1 - \alpha\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)\frac{\partial}{\partial w_j}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right) \\ &= w_1 - \alpha\sum_{n=1}^{5}\left(w_0 + w_1 x^{(n)} - y^{(n)}\right)x^{(n)}\end{aligned}$$

iii. Consider the initial parameters $\mathbf{w} = (0, 0)^{\mathrm{T}}$ and a learning rate $\alpha = 10^{-4}$. Update the parameters with the gradient descent rule. Plot the initial and the updated models in the same figure as the training set.

Solution:

$$w_0 = 0 - 10^{-4}\sum_{n=1}^{5}\left(-y^{(n)}\right) = 0 - 10^{-4}(-105) = 1.05 \cdot 10^{-2}$$

$$w_1 = 0 - 10^{-4}\sum_{n=1}^{5}\left(-y^{(n)}x^{(n)}\right) = 0 - 10^{-4}(-2775) = 0.2775$$

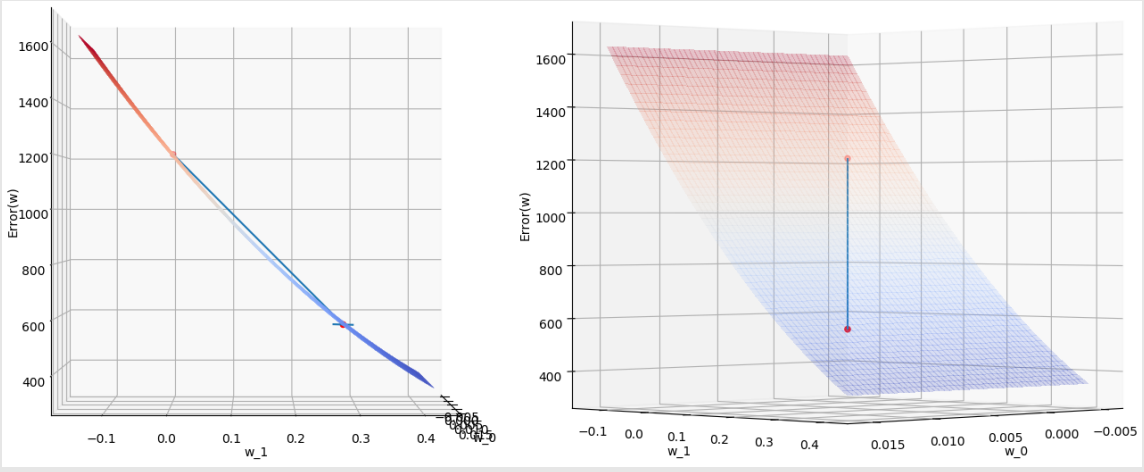


iv. Compute the error for the initial parameters and for the updated ones. Draw in 3D (axes $w_0, w_1, \mathbb{E}(w_0, w_1)$) the initial parameters and error, and the updated parameters with the current error.

Solution:

$$\mathbb{E}(\mathbf{w}_{\text{old}}) = \sum_{n=1}^{5} \left(w_0 + w_1 x^{(n)} - y^{(n)} \right)^2 = \sum_{n=1}^{5} {y^{(n)}}^2 = \frac{2775}{2} = 1212.5$$

$$\mathbb{E}(\mathbf{w}_{\text{new}}) = \sum_{n=1}^{5} \left(1.05 \cdot 10^{-2} + 0.2775 x^{(n)} - y^{(n)} \right)^2 = \cdots \approx 573.54$$



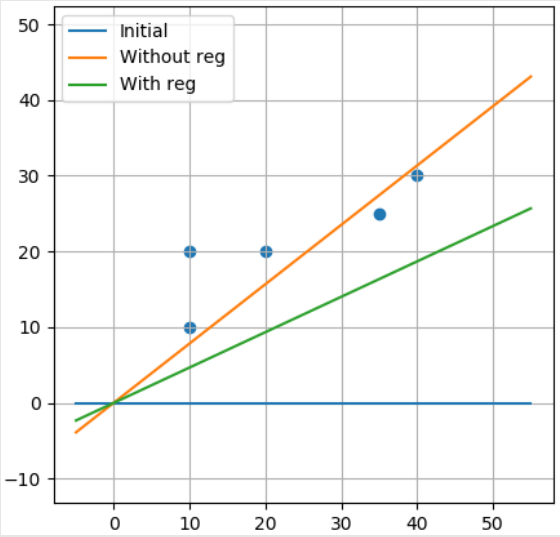
(c) *Gradient Descent (Jupyter Notebook)*: Estimate the parameters of a Linear Regression model using Gradient Descent (1) without regularisation, and (2) adding Ridge regression with $\lambda = 500$ (this value is too large, but consider it an experiment to see the effect of the regularisation).

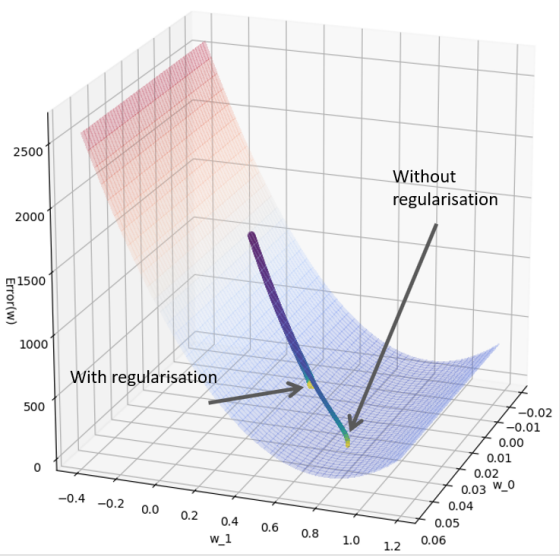
You may use the functions included in the Jupyter Notebook. Consider a tolerance of 10^{-6}, a learning rate of 10^{-4}, and a maximum number of iterations of 5000.

i. Compute the error in the training set for both models. Plot the training set and trained models in the same figure. Also, create another figure with the evolution of the parameters with respect to the error (3D plot). Use these two figures and the value of the errors to observe the effect of the

regularisation. Which conclusion do you extract? Reason your answer.

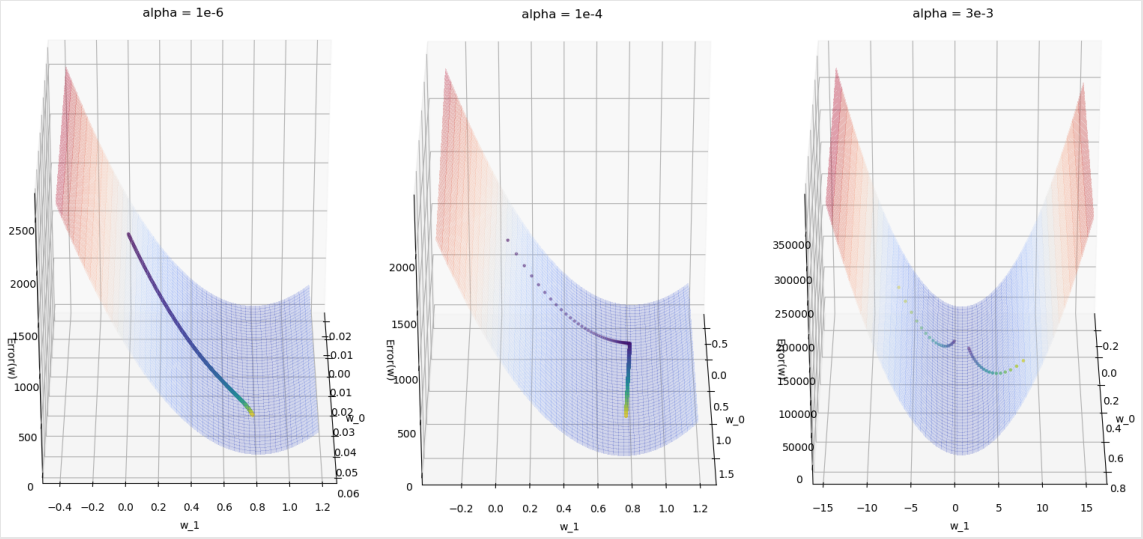
Solution:
Error without regularisation = 89.04
Error with regularisation = 289.34





ii. Try different values of the learning rate, $\alpha = 10^{-6}, 10^{-4}, 3 \cdot 10^{-3}$ and plot the evolution of the parameters with respect to the error for each of the cases. What is the difference between all the learning parameters? Reason your answer.

Solution:



2. *(Jupyter Notebook)* Consider the dataset given in the attached CSV file. Fit the best possible regression model to the given dataset using gradient descent. To do so, you will have to:

   - Choose the degree of the polynomial you want to fit.
   - Decide whether to use regularisation or not. In the positive case, write down the loss that is now minimised, and choose the value of the regularisation coefficient, λ.
   - Choose an optimal learning rate, α.

   Justify your choices and add figures to support them.

REVIEW: Vector and matrix derivatives

3. Let $A = \begin{pmatrix} 3 & 2 \\ 2 & 5 \end{pmatrix}$, and $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ calculate the expressions: $\mathbf{x}^T\mathbf{A}\mathbf{x}$, $\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial x_1}$, $\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial x_2}$ and $\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}}$. Compare the last result using the adequate rule of the table.

Solution: $\mathbf{x}^T\mathbf{A}\mathbf{x} = (x_1, x_2)\begin{pmatrix} 3 & 2 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2)\begin{pmatrix} 3x_1 + 2x_2 \\ 2x_1 + 5x_2 \end{pmatrix} = 3x_1^2 + 4x_1x_2 + 5x_2^2$

$\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial x_2} = 4x_1 + 10x_2$ $\quad$ $\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial x_1} = 6x_1 + 4x_2$, $\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \begin{pmatrix} 6x_1 + 4x_2 \\ 4x_1 + 10x_2 \end{pmatrix}$

If we use the formula: $\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} = 2\begin{pmatrix} 3 & 2 \\ 2 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6x_1 + 4x_2 \\ 4x_1 + 10x_2 \end{pmatrix}$

4. *(Optional)* Demonstrate the expressions:

   (a) $\frac{\partial \|\mathbf{x}\|^2}{\partial \mathbf{x}} = 2\mathbf{x}$ and $\quad \frac{\partial \|\mathbf{x}\|}{\partial \mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$.

   (b) $\mathbf{x}^T\mathbf{A}\mathbf{y} = \mathbf{y}^T\mathbf{A}^T\mathbf{x}$

   (c) Use the formulas of the table and the general operations of vectors and matrices to demonstrate the following expressions.

      i. Assume $\mathbf{A}$ is a symmetric matrix. Demonstrate $\frac{\partial (\mathbf{y} - \mathbf{x})^T\mathbf{A}(\mathbf{y} - \mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}(\mathbf{x} - \mathbf{y})$

      ii. Calculate $\frac{\partial \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2}{\partial \mathbf{x}} = 2\mathbf{A}^T(\mathbf{A} - \mathbf{b})$

Solution:

(a) $\|\mathbf{x}^T\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x} = (x_1, x_2)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1^2 + x_2^2$. $\frac{\partial \mathbf{x}^T\mathbf{x}}{\partial x_1} = 2x_1$ and $\frac{\partial \mathbf{x}^T\mathbf{x}}{\partial x_2} = 2x_2$. We have demonstrate that $\frac{\partial \mathbf{x}^T\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$

$$\frac{\partial \|\mathbf{x}\|}{\partial \mathbf{x}} = \frac{\partial \sqrt{\mathbf{x}^T\mathbf{x}}}{\partial \mathbf{x}} = \frac{\frac{\partial \mathbf{x}^T\mathbf{x}}{\partial \mathbf{x}}}{2\sqrt{\mathbf{x}^T\mathbf{x}}} = \frac{2\mathbf{x}}{2\|\mathbf{x}\|} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$$

(b) $\mathbf{x}^T\mathbf{A}\mathbf{y}$ is a number k.Given that the $k^T = k$: $\mathbf{x}^T\mathbf{A}\mathbf{y} = (\mathbf{x}^T\mathbf{A}\mathbf{y})^T = \mathbf{y}^T\mathbf{A}^T\mathbf{x}^{TT} = \mathbf{y}^T\mathbf{A}^T\mathbf{y}$.

We can also demonstrate the same developing each side of the formula:

$$\mathbf{x}^T\mathbf{A}\mathbf{y} = (x_1, x_2)\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = (x_1, x_2)\begin{pmatrix} a_{11}y_1 + a_{12}y_2 \\ a_{21}y_1 + a_{22}y_2 \end{pmatrix} = a_{11}x_1y_1 + a_{12}x_1y_2 + a_{21}x_2y_1 + a_{22}x_2y_2$$

$$\mathbf{y}^T\mathbf{A}^T\mathbf{x} = (y_1, y_2)\begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (y_1, y_2)\begin{pmatrix} a_{11}x_1 + a_{21}x_2 \\ a_{12}x_1 + a_{22}x_2 \end{pmatrix} = a_{11}y_1x_1 + a_{21}y_1x_2 + a_{12}y_2x_1 + a_{22}y_2x_2$$

(c) i. $(\mathbf{y} - \mathbf{x})^T\mathbf{A}(\mathbf{y} - \mathbf{x}) = (\mathbf{y}^T - \mathbf{x}^T)\mathbf{A}(\mathbf{y} - \mathbf{x}) = \mathbf{y}^T\mathbf{A}\mathbf{y} - \mathbf{y}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}\mathbf{y} + \mathbf{x}^T\mathbf{A}\mathbf{x} =$
note that $\mathbf{y}^T\mathbf{A}\mathbf{x}$ is a number, so $\mathbf{y}^T\mathbf{A}\mathbf{x} = (\mathbf{y}^T\mathbf{A}\mathbf{x})^T = \mathbf{x}^T\mathbf{A}^T\mathbf{y} = \mathbf{x}^T\mathbf{A}\mathbf{y}$ by the symmetry of $\mathbf{A}$
$= \mathbf{y}^T\mathbf{A}\mathbf{y} - \mathbf{x}^T\mathbf{A}\mathbf{y} - \mathbf{x}^T\mathbf{A}\mathbf{y} + \mathbf{x}^T\mathbf{A}\mathbf{x}$

$$\frac{\partial (\mathbf{y} - \mathbf{x})^T\mathbf{A}(\mathbf{y} - \mathbf{x})}{\partial \mathbf{x}} = 0 - \mathbf{A}\mathbf{y} - \mathbf{A}\mathbf{y} + 2\mathbf{A}\mathbf{x} = 2\mathbf{A}\mathbf{x} - 2\mathbf{A}\mathbf{y} = 2\mathbf{A}(\mathbf{x} - \mathbf{y})$$

ii. $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 = (\mathbf{A}\mathbf{x} - \mathbf{b})^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = (\mathbf{x}^T\mathbf{A}^T - \mathbf{b}^T)(\mathbf{A}\mathbf{x} - \mathbf{b}) = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{b}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{b} =$
$= \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} - \mathbf{x}^T\mathbf{A}^T\mathbf{b} + \mathbf{b}^T\mathbf{b}$

$$\frac{\partial \|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2}{\partial \mathbf{x}} = 2\mathbf{A}^T\mathbf{A}\mathbf{x} - \mathbf{A}^T\mathbf{b} - \mathbf{A}^T\mathbf{b} = 2\mathbf{A}^T\mathbf{A}\mathbf{x} - 2\mathbf{A}^T\mathbf{b} = 2\mathbf{A}^T(\mathbf{A} - \mathbf{b})$$

5. *(Optional)* Given $f(\mathbf{x}) = x_1^2 + x_2^2$ answer the following questions:

(a) Write it in the form $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}$.

(b) Plot the function f and the points $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ in the plane of the coordinate axis x_1x_2 where $f(\mathbf{x}) = 4$ (the *contour* of value 4 of f).

(c) Calculate and plot the gradient of f in the points where the contour intersect the axis x_1x_2 and also in the point $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

(d) (Generalisation)

i. Construct a polynomial function of second degree $g(\mathbf{x})$ if we know that the contour of $g(\mathbf{x}) = 1$ is an ellipse with the axis along the x_1 and x_2 axis and length of them 3 and 2 respectively.

ii. Calculate and draw the gradient in the points where the contour of $g(\mathbf{x}) = 4$ intersects the coordinate axis.

Solution:

(a) $f(\mathbf{x}) = (x_1, x_2)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$

(b) The contour are the points in the plane x_1x_2 and plot a circle or radius 2. The gradient $\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} = 2\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Then in the point $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ the vector gradient is $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$. In the points $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ -1 \end{pmatrix}$,are the vectors that result from changing the 1 by 2.

(c) i. $g(\mathbf{x}) = (x_1, x_2)\begin{pmatrix} 9 & 0 \\ 0 & 4 \end{pmatrix}^{-1}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (x_1, x_2)\begin{pmatrix} 1/9 & 0 \\ 0 & 1/4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{x_1^2}{9} + \frac{x_2^2}{4}$.

ii. The gradient of g will be:

$$\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} = 2\begin{pmatrix} 1/9 & 0 \\ 0 & 1/4 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} \frac{2x_1}{9} \\ \frac{x_2}{4} \end{pmatrix}$$

The contour $\frac{x_1^2}{9} + \frac{x_2^2}{4} = 4$ corresponds to the ellipse $\frac{x_1^2}{36} + \frac{x_2^2}{16} = 1$ with axis length of 6 and 4. The point $\begin{pmatrix} 6 \\ 0 \end{pmatrix}$ has the gradient vector $\begin{pmatrix} 4/3 \\ 0 \end{pmatrix}$ and similarly for the other points $\begin{pmatrix} 0 \\ 4 \end{pmatrix}, \begin{pmatrix} -6 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ -4 \end{pmatrix}$.