DSC 40A Fall 2025 - Group Work Session 4
due Monday, October 20th at 11:59PM

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. **One person** from each group should submit your solutions to Gradescope and **tag all group members** so everyone gets credit.

This worksheet won't be graded on correctness, but rather on good-faith effort. Even if you don't solve any of the problems, you should include some explanation of what you thought about and discussed, so that you can get credit for spending time on the assignment.

In order to receive full credit, you must work in a group of two to four students for at least 50 minutes in your assigned discussion section. You can also self-organize a group and meet outside of discussion section for 80 percent credit. You may not do the groupwork alone.

Note(!!): this groupwork is slightly longer. You should plan on completing problems 1 through 5 in about 50 minutes. After working on the worksheet for 50 minutes, submit what you have completed.

1 Gradient of Scalar-valued Functions

As we start dabbling in multiple linear regression and linear algebra, we will start to work with vector-valued functions. One example of a vector-valued function we have seen in this class is the empirical risk for linear regression.

$$R_{\text{sq}}(\vec{w}) = R_{\text{sq}}\left(\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}\right) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (w_0 + w_1 x_i))^2$$

Recall that to solve for the optimal model parameters, w_0^* and w_1^*, we found the partial derivatives of the empirical risk with respect to each parameter individually. Observe that these make up the components of the gradient of the empirical risk function:

$$\nabla R_{\text{sq}}(\vec{w}) = \begin{bmatrix} \frac{\partial R_{\text{sq}}}{\partial w_0} \\ \frac{\partial R_{\text{sq}}}{\partial w_1} \end{bmatrix}$$

In essence, **the gradient of a vector-valued function is a vector of the function's partial derivatives with respect to each component of the vector**. Notice that this means the dimensions of the function's domain should match the dimensions of the gradient. With regards to the above empirical risk example, R_{sq} takes in a vector of length 2, namely $\begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$, so its gradient is also a vector with 2 components. Keep this in mind when solving for gradients.

Problem 1.

Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is the function:

$$f(\vec{x}) = 3x_1^2 + 2x_1x_2 - x_1\cos(x_3)$$

What is the gradient of f? *Hint: Start by finding the partial derivatives of f with respect to each of x_1, x_2, and x_3.*

Solution:

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} 6x_1 + 2x_2 - \cos(x_3) \\ 2x_1 \\ x_1 \sin(x_3) \end{bmatrix}$$

Problem 2.

Suppose $\vec{a} \in \mathbb{R}^3$, and suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is the function:

$$f(\vec{x}) = \vec{a}^T \vec{x} = a_1 x_1 + a_2 x_2 + a_3 x_3$$

What is the gradient of f? *Hint: Again, start by finding the partial derivatives of f with respect to each of x_1, x_2, and x_3.*

Solution:

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \vec{a}$$

Problem 3.

Perhaps you have begun to notice that some derivative tricks you learned in single variable calculus also apply with vector calculus. We will make some connections to rules you saw in single variable calculus, like the product rule, in this problem.

a) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is the function:

$$f(\vec{x}) = x_1^2 + x_2^2 + ... + x_n^2$$

What is the gradient of f?

Solution:

$$\nabla f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{bmatrix} = 2 \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = 2\vec{x}$$

b) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is the function:

$$f(\vec{x}) = \vec{x}^T \vec{x}$$

Using the result from Problem 3 and the product rule for single variable functions, what is the gradient of f? *Hint: Recall the product rule for single variable functions, $\frac{d}{dx} f(x)g(x) = f'(x)g(x) + f(x)g'(x)$.*

Solution:

$$\nabla f(\vec{x}) = \vec{x} + \vec{x} = 2\vec{x}$$

c) Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is the function:

$$f(\vec{x}) = \|\vec{x}\|^2$$

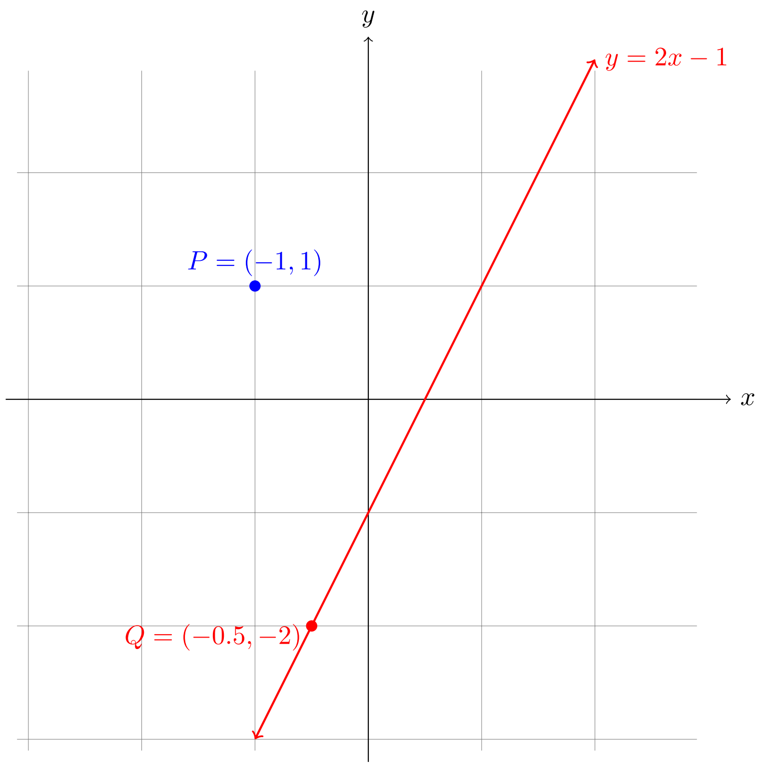
What is the gradient of f?

Solution: $$\nabla f(\vec{x}) = 2\vec{x}$$

Note that all three functions f above are really the same function!

2 Fun with Geometry - Draw a picture

The purpose of this problem is to help build (or re-build) our skills and intuition for the rich connections between geometry and linear algebra.



Problem 4.

In this problem we will use the drawing above and some linear algebra to find the distance between the point $P(-1,1)$ and the red line $y = 2x + 1$.

a) A useful fact that might be familiar is that all lines in the (x, y)-plane can be written in the form

$$L = \left\{ \mathbf{x} = (x, y) : \mathbf{x}^T \mathbf{n} = c \right\}$$

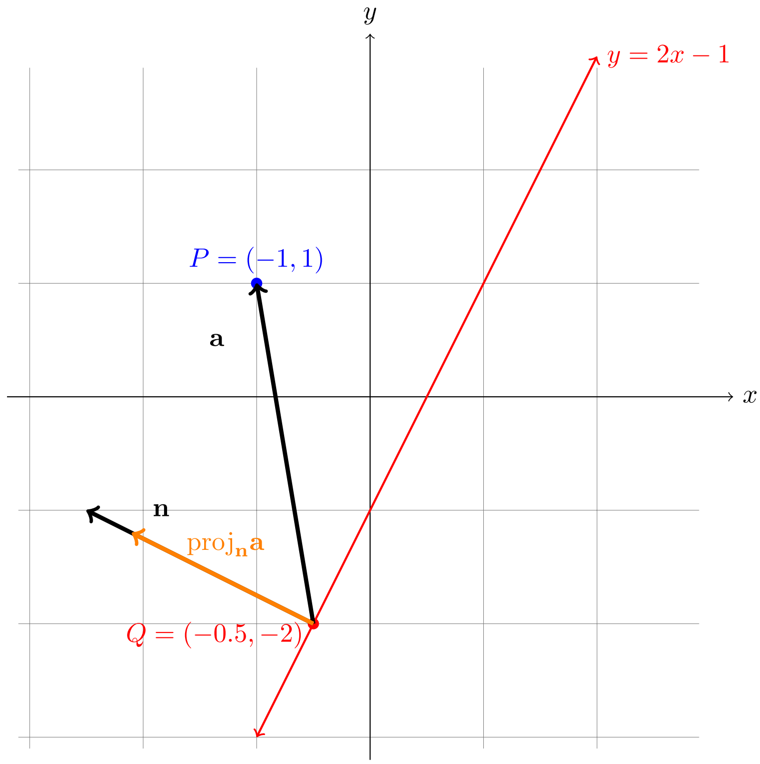
for some vector $\mathbf{n}$ called the **normal vector** and a scalar $c \in \mathbb{R}$ called the **offset**. Using the equation $y = 2x + 1$ that defines the red line, manipulate the equation to show that $\mathbf{n} = (-2, 1)$ is a normal vector for the line. Then, find the offset c.

Solution: $y = 2x + 1$ is equivalent to $-2x + y = 1$, or $\mathbf{x}^T(-2, 1) = 1$. Therefore $\mathbf{n} = (-2, 1)$ and $c = 1$.

b) In the picture above, draw the following three vectors **originating from point Q (!!)**:

1. The vector $\mathbf{a} = \overrightarrow{QP}$,
2. The normal vector $\mathbf{n}$ from part **(a)**.
3. The orthogonal projection $\mathbf{p} = \text{proj}_{\mathbf{n}} \mathbf{a}$. You will need to calculate it first.

Solution:



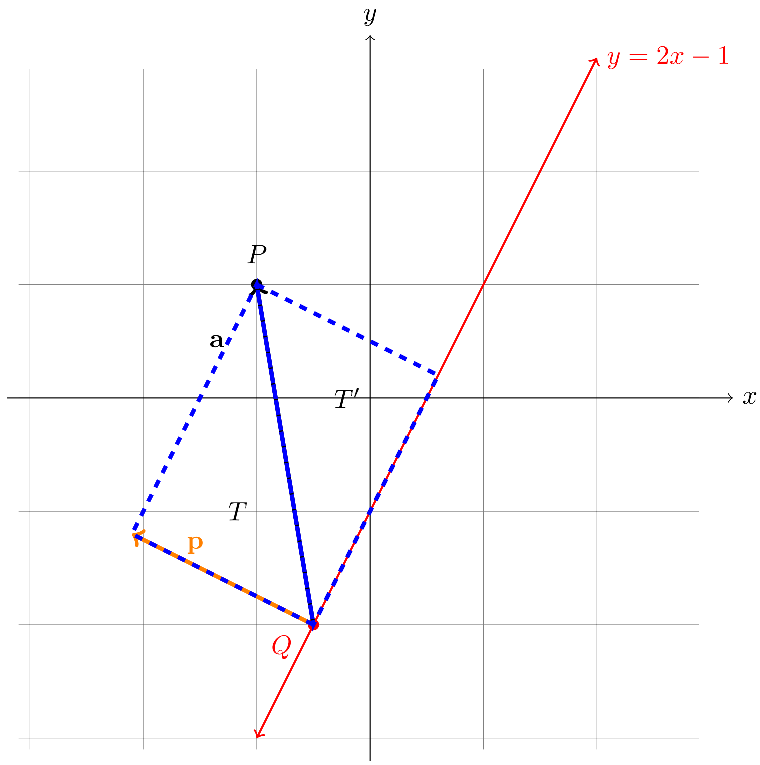
To find $\mathbf{p}$, start by writing $\mathbf{a} = P - Q = (-0.5, 3)$. Then

$$\mathbf{p} = \frac{\mathbf{n}^T\mathbf{a}}{\mathbf{n}^T\mathbf{n}}\mathbf{n} \tag{1}$$

$$= \frac{1+3}{4+1}(-2, 1) = (-1.6, 0.8) \tag{2}$$

c) Using all of these ingredients, find the distance between P and the red line. *HINT: Draw a right triangle in the picture with vertices P, Q, and the endpoint of the projection vector* $\mathbf{p}$.

Solution: Based on the hint, we have the picture below. Notice that the triangle T described in the hint is congruent to a triangle T' reflected across the vector $\overrightarrow{QP}$.



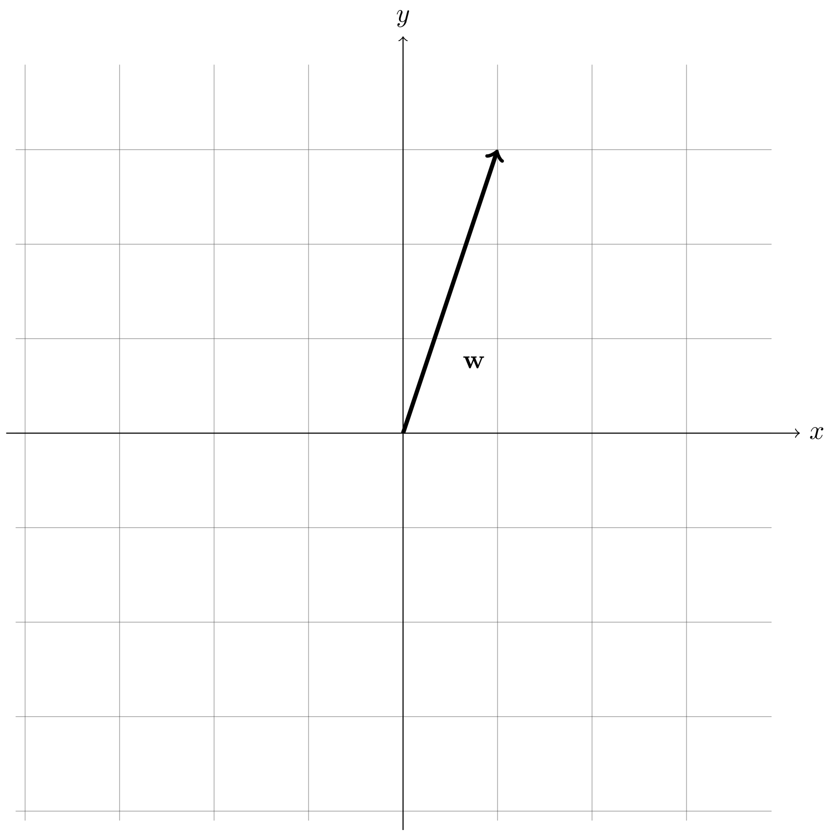
Therefore, the distance between P and the line is given by the length of the side of the triangle T on the bottom left, which is $\|\mathbf{p}\|$. We therefore have

$$\text{Distance} = \|\mathbf{p}\| = \sqrt{1.6^2 + 0.8^2} \approx 1.788.$$

3 Support Vector Machine v0.0.1

The Linear Support Vector Machine is a popular machine learning model. **But how does it work?** One's journey in becoming a successful data scientist is full of questions like these.

In the next two problems we will build a simplified version of the support vector machine model on a toy dataset. We will use some tools and skills that we explored in Section 1 - please make sure you have thoroughly explored it.



Problem 5.

In the illustration above we have rendered a vector $\mathbf{w} = (1, 3)$.

a) Consider the set $Z = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x}^T\mathbf{w} = 0\}$. Show that Z is a vector subspace of $\mathbb{R}^2$ and then draw it in the picture above.

Solution: $\mathbf{x}^T\mathbf{w} = 0$ implies that $x_1w_1 + x_2w_2 = 0$. In this case, $x_1 + 3x_2 = 0$. To show that Z is a vector subspace of $\mathbb{R}^2$, we need to show:

1. Contains the zero vector.

Substitute the zero vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ into the equation we get $0 + 3(0) = 0$.

2. Closed under addition.

Let $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\vec{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in Z$.This implies $x_1 + 3x_2 = 0$ and $y_1 + 3y_2 = 0$, and

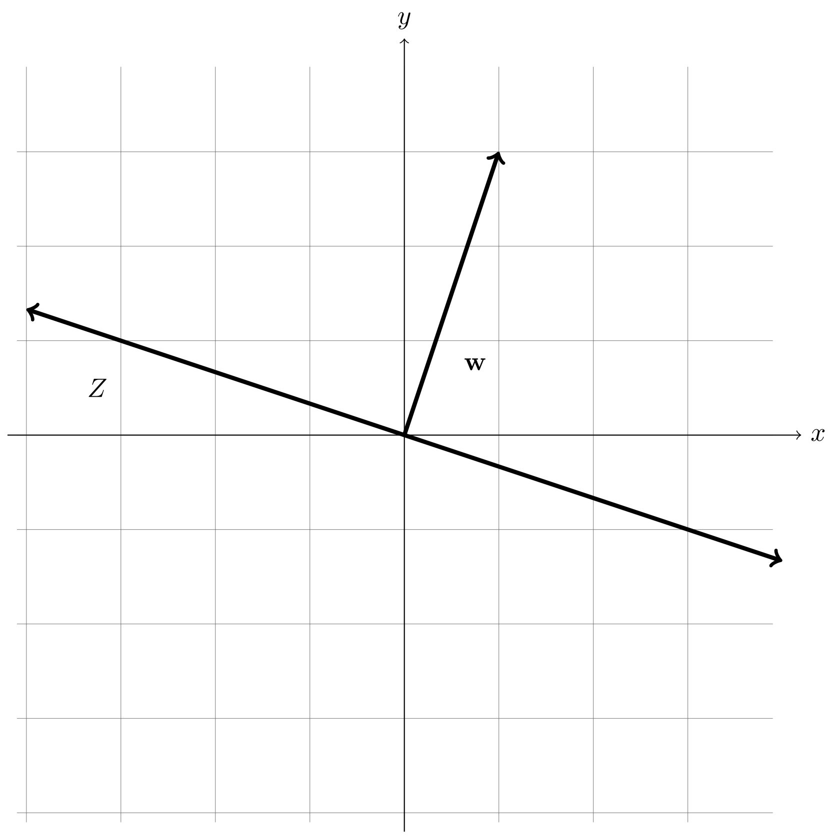
$$\vec{x} + \vec{y} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix}$$

$$x_1 + y_1 + 3(x_2 + y_2) = (x_1 + 3x_2) + (y_1 + 3y_2) = 0 + 0 = 0.$$

3. Closed under scalar multiplication.
Let $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and let $\lambda \in \mathbf{R}$.

$$\lambda\vec{x} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix}$$

$$\lambda x_1 + \lambda x_2 = \lambda x_1 + \lambda \cdot 3x_2 = \lambda(x_1 + 3x_2) = \lambda \cdot 0 = 0.$$

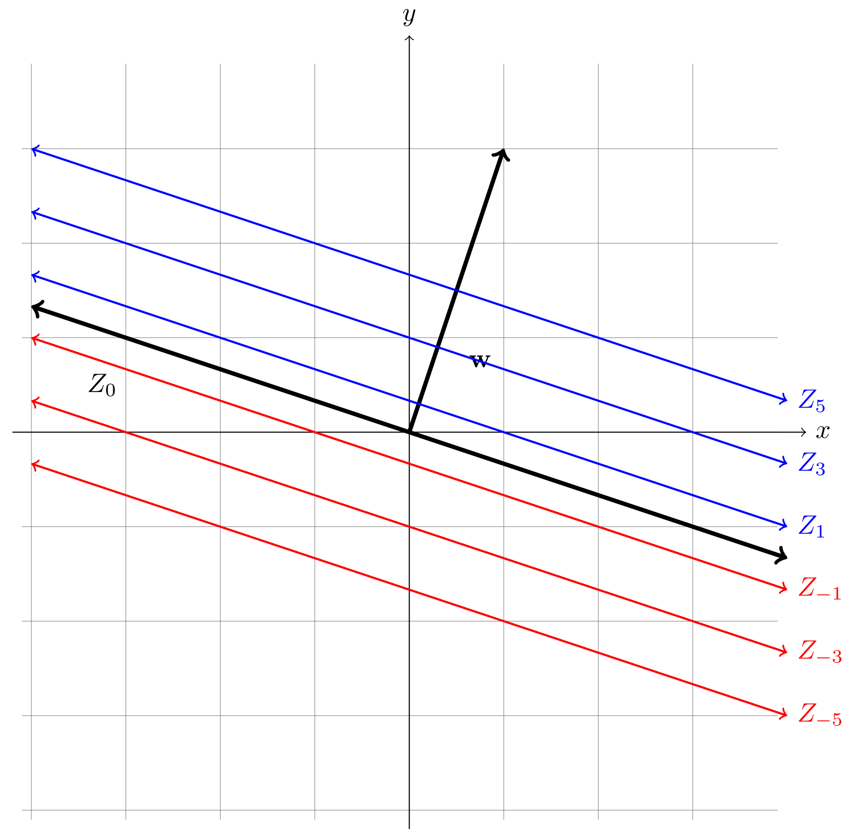


b) Now let a scalar $c \in \mathbb{R}$ be fixed. Let $Z_c = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x}^T\mathbf{w} = c\}$. Is Z_c a subspace? Why or why not? Draw Z_c in your picture for $c = -5, -3, -1, 1, 3, 5$. Note: You have already drawn Z_0 in the previous part. Once you get the idea of how these lines are situated you can sketch your informed guesses - no need to be 100% exact here.

Solution: Z_c is a subspace only if $c = 0$. The given information implies $x_1 + 3x_2 = c$.
1. Plug in the zero vector we get $0 + 3(0) = c$. We will only get the zero vector if $c = 0$.
2. To have this space closed under addition, we need to have that

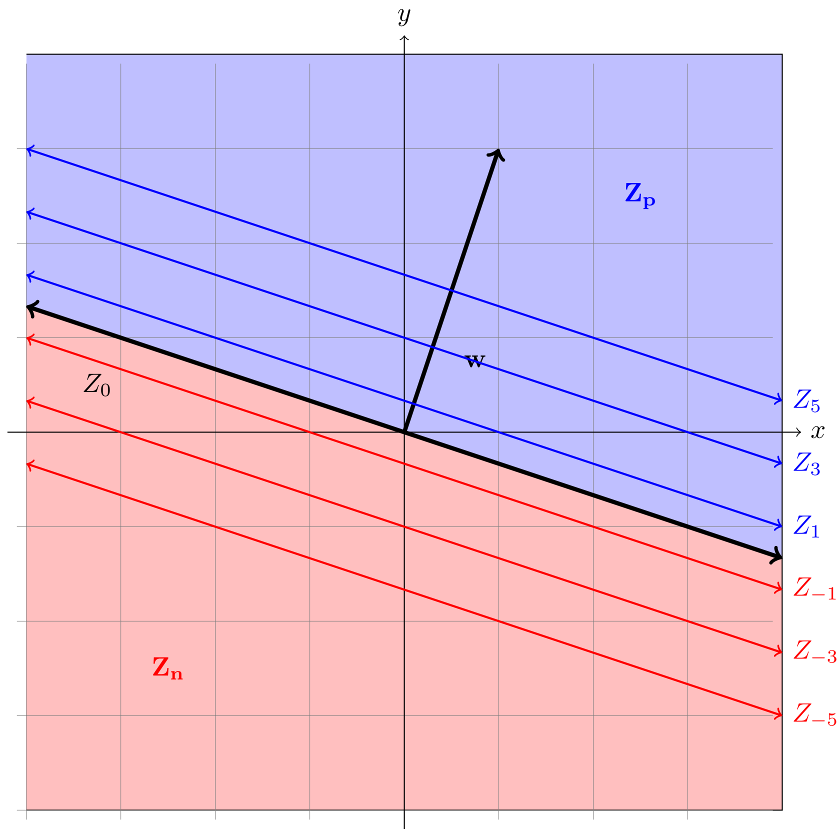
$$(x_1 + y_1) + 3(x_2 + y_2) = (x_1 + 3x_2) + (y_1 + 3y_2) = c + c = 2c = 0$$

Again showing c = 0 needs to be satisfied.
3. Let $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in Z_c$, meaning $x_1 + 3x_2 = c$. Let $\lambda \in \mathbb{R}$. We need it so that $\lambda(x_1 + 3x_2) = 0$. The only way this can be guaranteed is when $c = 0$.



c) As a final ingredient, let $Z_p = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x}^T\mathbf{w} > 0\}$ and let $Z_n = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x}^T\mathbf{w} < 0\}$. What do these regions look like? Shade them in the picture.

Solution:



Problem 6.

In this problem we will introduce the Simplified Linear SVM (SLSVM) and build one based on an example dataset. Linear support vector machines are often used for **binary classification tasks**, which are situations where your data falls into one of two classes and in which you wish to design a model to predict the class of incoming data. Some examples: identifying whether an image contains a cat or a dog, whether a leaf is healthy or diseased, and so on.

Assume that your dataset is of the form $(\mathbf{x}^{(i)}, y^{(i)})$ where each $\mathbf{x}^{(i)} \in \mathbb{R}^2$ and each $y^{(i)} \in \{+1, -1\}$ (+1 and -1 correspond to the class labels).

a) Consider the hypothesis function $H(\mathbf{x}^{(i)}) = \mathbf{w}^T\mathbf{x}$. What are the parameters of this model? Then, using your intuition from the previous problem, **describe geometrically** how the hypothesis H is attempting to classify your data.

> **Solution:** The parameters are $\mathbf{w} = (w_1, w_2)$. Based on the previous problem, the model is separating the plane into two regions based on either side of the line $\mathbf{w}^T\mathbf{x} = 0$. Then making a class prediction based on which side of this line your input data falls.

b) For the dataset $(\mathbf{x}^{(i)}, y^{(i)})$, write the empirical loss functions R_{abs} and R_{sq} for the absolute and square losses, respectively.

> **Solution:**
>
> $$R_{\text{abs}} = \frac{1}{n}\sum_{i=1}^{n}\left|y^{(i)} - H(x^*)\right|$$

$$R_{\text{sq}} = \frac{1}{n}\sum_{i=1}^{n}\left(y^{(i)} - H(x^*)\right)^2$$

c) Observe that the gradient gradient $\nabla_{\mathbf{w}} R_{\text{sq}}$ is given by the formula

$$\nabla_{\mathbf{w}} R_{\text{sq}} = \left\langle \frac{\partial}{\partial w_1} R_{\text{sq}}(\mathbf{x}^{(i)}, y^{(i)}), \frac{\partial}{\partial w_2} R_{\text{sq}}(\mathbf{x}^{(i)}, y^{(i)}) \right\rangle \tag{3}$$

$$= \left\langle \frac{2}{n}\sum_{i=1}^{n} y^{(i)}\mathbf{x}_1^{(i)} - \mathbf{w}^T\left(\frac{2}{n}\sum_{i=1}^{n}\mathbf{x}_1^{(i)}\mathbf{x}^{(i)}\right), \frac{2}{n}\sum_{i=1}^{n} y^{(i)}\mathbf{x}_2^{(i)} - \mathbf{w}^T\left(\frac{2}{n}\sum_{i=1}^{n}\mathbf{x}_2^{(i)}\mathbf{x}^{(i)}\right)\right\rangle \tag{4}$$

You do not need to prove this during discussion! We recommend trying to prove this formula on your own time. Although it looks pretty terrifying, observe that it is actually a simple structure if we hide some of the notation (this is a theme in mathematics). Let $\alpha_1 = \frac{2}{n}\sum_{i=1}^{n} y^{(i)}\mathbf{x}_1^{(i)}$ and $\alpha_2 = \frac{2}{n}\sum_{i=1}^{n} y^{(i)}\mathbf{x}_2^{(i)}$. Then, let $\mathbf{a}_1 = \frac{2}{n}\sum_{i=1}^{n}\mathbf{x}_1^{(i)}\mathbf{x}^{(i)}$ and $\mathbf{a}_2 = \frac{2}{n}\sum_{i=1}^{n}\mathbf{x}_2^{(i)}\mathbf{x}^{(i)}$. Then we have

$$\nabla_{\mathbf{w}} R_{\text{sq}} = \left\langle \alpha_1 - \mathbf{w}^T\mathbf{a_1}, \alpha_2 - \mathbf{w}^T\mathbf{a_2}\right\rangle = 0. \tag{5}$$

Thoroughly investigate these and convince yourself that α_1, α_2 are actually scalars and $\mathbf{a}_1, \mathbf{a}_2$ are actually vectors.

Then, show that any vector $\mathbf{w}^*$ which satisfies $\nabla R_{\text{sq}}(\mathbf{w}^*) = \mathbf{0}$ must satisfy the following equation:

$$\begin{bmatrix}\mathbf{a}_1^T \\ \mathbf{a}_2^T\end{bmatrix}\begin{bmatrix}w_1 \\ w_2\end{bmatrix} = \begin{bmatrix}\alpha_1 \\ \alpha_2\end{bmatrix} \tag{6}$$

Solution: Starting with the gradient equation, we have

$$\nabla_{\mathbf{w}} R_{\text{sq}} = \left\langle \alpha_1 - \mathbf{w}^T\mathbf{a_1}, \alpha_2 - \mathbf{w}^T\mathbf{a_2}\right\rangle = 0 \tag{7}$$

$$\begin{cases}\alpha_1 - \mathbf{w}^T\mathbf{a_1} & = 0 \\ \alpha_2 - \mathbf{w}^T\mathbf{a_2} & = 0\end{cases} \tag{8}$$

Therefore we must have $\mathbf{w}^T\mathbf{a_1} = \alpha_1$ and $\alpha_2 - \mathbf{w}^T\mathbf{a_2} = \alpha_2$. This is the same as writing

$$\begin{bmatrix}\mathbf{a}_1^T \\ \mathbf{a}_2^T\end{bmatrix}\begin{bmatrix}w_1 \\ w_2\end{bmatrix} = \begin{bmatrix}\alpha_1 \\ \alpha_2\end{bmatrix} \tag{9}$$

d) Now consider the following example. Robert is a biologist at UC Irvine and studies minnows. Some of the fish in his lab are infected with a disease *Infectus Horriblus*, and some are healthy. For four fish, he records their length in cm and their weight in g and records them in the table below. He then observes whether they are infected or healthy.

Length (cm)	Weight (g)	Diseased
1.0	2	Healthy
1.1	1.85	Healthy
1.75	0.9	Diseased
2.9	1.1	Diseased

Using part (d), find $\alpha_1, \alpha_2, \mathbf{a}_1, \mathbf{a}_2$ and write down the linear equation for the optimal SLSVM for this dataset. You will need to use a calculator. (Let $y_i = 1$ if the minnow is healthy and $y_i = -1$ if it is diseased.)

Solution: Let $y_i = 1$ if the minnow is healthy and $y_i = -1$ if it is diseased. Then we calculate:

$$\alpha_1 = \frac{2}{n}\sum_{i=1}^{n} y^{(i)}\mathbf{x}_1^{(i)} \tag{10}$$
$$= \frac{2}{4}\left((1)(1.0) + (1)(1.1) + (-1)(1.75) + (-1)(2.9)\right) = -1.275 \tag{11}$$
$$\alpha_2 = \frac{2}{n}\sum_{i=1}^{n} y^{(i)}\mathbf{x}_2^{(i)} \tag{12}$$
$$= \frac{2}{4}\left((1)(2.0) + (1)(1.85) + (-1)(0.9) + (-1)(1.1)\right) = 0.925 \tag{13}$$
$$\mathbf{a}_1 = \frac{2}{n}\sum_{i=1}^{n} \mathbf{x}_1^{(i)}\mathbf{x}^{(i)} \tag{14}$$
$$= \frac{2}{4}\left(1.0(1,2) + 1.1(1.1,1.85) + 1.75(1.75,0.9) + 2.9(2.9,1.1)\right) = (6.84125, 4.4) \tag{15}$$
$$\mathbf{a}_2 = \frac{2}{n}\sum_{i=1}^{n} \mathbf{x}_2^{(i)}\mathbf{x}^{(i)} \tag{16}$$
$$= \frac{2}{4}\left(2.0(1,2) + 1.85(1.1,1.85) + 0.9(1.75,0.9) + 1.1(2.9,1.1)\right) = (4.4, 4.72125) \tag{17}$$
$$\tag{18}$$

Therefore the linear system is given by

$$\begin{bmatrix} 6.84125 & 4.4 \\ 4.4 & 4.72125 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} -1.275 \\ 0.925 \end{bmatrix} \tag{19}$$

e) Use an online calculator to solve this sytem and show that

$$\mathbf{w}^* \approx \begin{bmatrix} -0.78 \\ 0.92 \end{bmatrix} \tag{20}$$

On Desmos, plot your original data and the line $\{\mathbf{x}^T\mathbf{w}^* = 0\}$, which is called the **decision boundary**.

Solution: