DSC 140A - Homework 03

Due: Wednesday, January 31

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope at 11:59 PM.

Problem 1.

Suppose that $f : \mathbb{R}^d \to \mathbb{R}$ is convex and $g : \mathbb{R} \to \mathbb{R}$ is convex and non-decreasing. That is, if $a > b$, then $g(a) \geq g(b)$.

Show that the composition of these functions, $h(\vec{x}) = g(f(\vec{x}))$, is also convex.

Hint: you'll want to go back to the definition to show this is true. A similar problem was solved in discussion.

Solution:

We need to show that, for any $\vec{a}, \vec{b} \in \mathbb{R}^d$ and $t \in [0, 1]$, the following holds:

$$h(t\vec{a} + (1-t)\vec{b}) \leq th(\vec{a}) + (1-t)h(\vec{b}).$$

Starting with the left-hand side and using the definition of h, we have:

$$h(t\vec{a} + (1-t)\vec{b}) = g(f(t\vec{a} + (1-t)\vec{b}))$$

Because f is convex, we know that $f(t\vec{a} + (1-t)\vec{b}) \leq tf(\vec{a}) + (1-t)f(\vec{b})$. Combining this with the fact that g is non-decreasing, we have:

$$\leq g(tf(\vec{a}) + (1-t)f(\vec{b}))$$

Now we use the fact that g is convex to write:

$$\leq tg(f(\vec{a})) + (1-t)g(f(\vec{b}))$$

Which we recognize as the right-hand side of the inequality we want to show:

$$= th(\vec{a}) + (1-t)h(\vec{b}).$$

Problem 2.

Recall that the *hinge loss* is

$$L_{\text{hinge}}(\vec{w}, \vec{x}, y) = \max\{0, 1 - y\,\vec{w} \cdot \text{Aug}(\vec{x})\}$$

The Soft-SVM problem aims to minimize the regularized empirical risk:

$$R(\vec{w}) = \frac{C}{n} \sum_{i=1}^{n} L_{\text{hinge}}(\vec{w}, \vec{x}^{(i)}, y_i) + \|\vec{w}\|^2$$

Show that $R(\vec{w})$ is a convex function of $\vec{w}$.

Hint: you will probably *not* want to use the formal definition of convexity here. Instead, you'll want to show that R is composed of simpler functions which themselves are convex.

Solution: We will break the function into pieces and show that each piece is convex, and so the sum of the pieces is also convex.

First, $\|\vec{w}\|^2$ is a convex function of $\vec{w}$. This can be shown in several ways; the simplest might be the second derivative test. Remember that $\|\vec{w}\|^2 = \vec{w} \cdot \vec{w} = w_1^2 + w_2^2 + \ldots + w_d^2$. The first derivative $\partial / \partial w_i$ is $2w_i$, and the second derivative $\partial / \partial w_i \partial w_j$ is 2 if $i = j$ and 0 otherwise. Therefore, the Hessian matrix is:

$$\begin{pmatrix} 2 & 0 & \cdots & 0 \\ 0 & 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 2 \end{pmatrix} = 2I,$$

where I is the identity matrix. This matrix is positive semi-definite, so $\|\vec{w}\|^2$ is convex.

To show that $C/n \sum_{i=1}^{n} L_{\text{hinge}}(\vec{w}, \vec{x}^{(i)}, y_i)$ is convex, we will show that the hinge loss is convex and then use the fact that the sum of convex functions is convex.

The hinge loss can be written as the maximum of two functions:

$$L_{\text{hinge}}(\vec{w}, \vec{x}, y) = \max\{0, 1 - y\,\vec{w} \cdot \text{Aug}(\vec{x})\}.$$

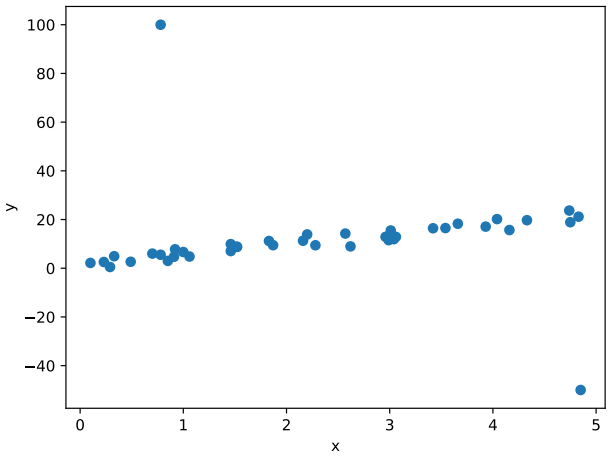
From lecture, we know that the maximum of two convex functions is convex, so we only need to show that each of the two functions is convex. The first function, 0, is a constant function and is therefore convex. The second function, $1 - y\,\vec{w} \cdot \text{Aug}(\vec{x})$, is also convex. We can most easily show this using the result from Discussion, which says that functions of the form $a\vec{x} \cdot \vec{w} - b$ are convex functions of $\vec{w}$.

To show that $1 - y\,\vec{w} \cdot \text{Aug}(\vec{x})$ is of this form, we set $a = -y$ and $b = -1$. Therefore, the function is convex, the hinge loss is convex, and the sum of hinge losses is convex.

Lastly, since $C/n \geq 0$, the first term is convex, and so the whole thing is convex.

Problem 3.

In the last homework, you fit a line to the data set below by minimizing the mean squared error:



Notably, the data contains outliers which may affect our regression.

Remember that you can download the data set at the link below:

`https://f000.backblazeb2.com/file/jeldridge-data/002-regression_outlier/data.csv`

a) Recall the absolute loss:

$$L_{\text{abs}}(H(x), y) = |H(x) - y|.$$

The absolute loss is not as sensitive to outliers as compared to the square loss. However, it is not differentiable, so unlike the mean squared error there is no closed form solution for the minimizer of the risk.

Implement subgradient descent and use it to fit a function of the form $w_0 + w_1 x$ by minimizing the risk with respect to the absolute loss. Report w_0 and w_1, and provide your code.

Hint: the subgradient of the absolute loss was considered in discussion.

Solution: Recall that a subgradient of the absolute loss is:

$$g(\vec{w}, \vec{x}, y) = \text{sign}(\vec{x} \cdot \vec{w} - y)\vec{x}$$

With this information:

```
def subgradient_descent(x, subgradient, learning_rate, tol=1e-5):
    i = 1
    while True:
        g = subgradient(x)
        eta = learning_rate(i)

        if np.linalg.norm(eta * g) < tol:
            break

        x = x - eta * g
        i += 1
    return x

def g_l(w, x, y):
    """Subgradient of the absolute loss w.r.t. w"""
    return np.sign(x @ w - y) * x

def subgradient(w):
    """Subgradient of the risk."""
    return np.mean([g_l(w, x, y) for x, y in zip(X, y)], axis=0)

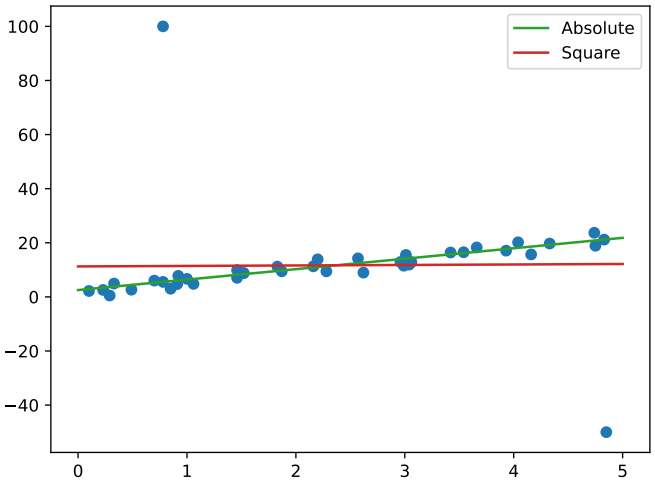
def learning_rate(n):
    return 1 / np.sqrt(n)
```

Running this, we converge at weights of: $w_0 \approx 2.53$ and $w_1 \approx 3.86$.

b) Plot both of the fitted lines on top of a scatter plot of the data; that is, plot the least squares regression line you found in Homework 02 and the line fit by minimizing the risk of the absolute loss.

You should see that the two lines are quite different. Briefly explain why this is the case.

Solution:



The slope of the line fit by minimizing the square loss has been affected by the outliers; the point at (1, 100) is "pulling" the line upwards, towards it. The slope of the line fit by minimizing the absolute loss is not as affected by the outliers.