MAD - assignment 1

Frederik Altmann

November 2024

1 Exercise 1

Partial Derivatives

1.1 a)

$$
\begin{aligned}
f(x,y) &= x^4y^3 + 7x^5 - e^{xy} \\
\frac{\partial f}{\partial x} &= 4x^3y + 5 \cdot 7x^4 - ye^{xy} \\
\frac{\partial f}{\partial y} &= 3x^4y^2 - xe^{xy}
\end{aligned}
$$

1.2 b)

$$
\begin{aligned}
f(x,y) &= \frac{1}{\sqrt{x^3+xy+y^2}} \\
f(x,y) &= (x^3+xy+y^2)^{-\frac{1}{2}} \\
\frac{\partial f}{\partial x} &= -\frac{1}{2}(x^3+xy+y^2)^{-\frac{3}{2}} \cdot \frac{\partial}{\partial x}(x^3+xy+y^2) \\
\frac{\partial}{\partial x}(x^3+xy+y^2) &= \frac{\partial}{\partial x}(x^3) + \frac{\partial}{\partial x}(xy) + \frac{\partial}{\partial x}(y^2) \\
&= 3x^2 + y + 0.
\end{aligned}
$$

Substitute:

$$
\begin{aligned}
\frac{\partial f}{\partial x} &= -\frac{1}{2}(x^3+xy+y^2)^{-\frac{3}{2}} \cdot (3x^2+y) \\
\frac{\partial f}{\partial x} &= -\frac{1}{2} \cdot \frac{3x^2+y}{(x^3+xy+y^2)^{\frac{3}{2}}}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f}{\partial y} &= -\frac{1}{2}(x^3+xy+y^2)^{-\frac{3}{2}} \cdot \frac{\partial}{\partial y}(x^3+xy+y^2) \\
\frac{\partial}{\partial y}(x^3+xy+y^2) &= \frac{\partial}{\partial y}(x^3) + \frac{\partial}{\partial y}(xy) + \frac{\partial}{\partial y}(y^2) \\
&= 0 + x + 2y.
\end{aligned}
$$

Substitute:

$$
\begin{aligned}
\frac{\partial f}{\partial y} &= -\frac{1}{2}(x^3+xy+y^2)^{-\frac{3}{2}} \cdot (x+2y) \\
\frac{\partial f}{\partial y} &= -\frac{1}{2} \cdot \frac{x+2y}{(x^3+xy+y^2)^{\frac{3}{2}}}
\end{aligned}
$$

1.3 c)

$$
\begin{aligned}
f(x,y) &= \frac{x^3+y^2}{x+y}. \\
\frac{\partial f}{\partial x} &= \frac{\partial}{\partial x}\left(\frac{x^3+y^2}{x+y}\right) \\
\frac{\partial}{\partial x}\left(\frac{u}{v}\right) &= \frac{v\frac{\partial u}{\partial x} - u\frac{\partial v}{\partial x}}{v^2} \\
u &= x^3+y^2 \quad \text{and} \quad v = x+y \\
\frac{\partial u}{\partial x}: \quad \frac{\partial}{\partial x}(x^3+y^2) &= 3x^2+0 = 3x^2 \\
\frac{\partial v}{\partial x}: \quad \frac{\partial}{\partial x}(x+y) &= 1 \\
\text{Substitute} \quad \frac{\partial f}{\partial x} &= \frac{(x+y)(3x^2)-(x^3+y^2)(1)}{(x+y)^2}. \\
&= \frac{3x^2(x+y)-(x^3+y^2)}{(x+y)^2}. \\
&= \frac{3x^3+3x^2y-x^3-y^2}{(x+y)^2}. \\
&= \frac{2x^3+3x^2y-y^2}{(x+y)^2}.
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial f}{\partial y} &= \frac{\partial}{\partial y}\left(\frac{x^3+y^2}{x+y}\right) \\
\frac{\partial}{\partial y}\left(\frac{u}{v}\right) &= \frac{v\frac{\partial u}{\partial y} - u\frac{\partial v}{\partial y}}{v^2} \\
u &= x^3+y^2 \quad \text{and} \quad v = x+y \\
\frac{\partial u}{\partial y}: \quad \frac{\partial}{\partial y}(x^3+y^2) &= 0+2y = 2y \\
\frac{\partial v}{\partial y}: \quad \frac{\partial}{\partial y}(x+y) &= 1 \\
\frac{\partial f}{\partial y} &= \frac{(x+y)(2y)-(x^3+y^2)(1)}{(x+y)^2} \\
&= \frac{2y(x+y)-(x^3+y^2)}{(x+y)^2} \\
&= \frac{2xy+2y^2-x^3-y^2}{(x+y)^2} \\
&= \frac{-x^3+2xy+y^2}{(x+y)^2}
\end{aligned}
$$

2 Exercise 2

Gradients

2.1 a

$$
\begin{aligned}
f(\bar{x}) &= \bar{x}^T\bar{x} + c \\
&\quad \text{Gradient of } \bar{x}^T\bar{x} \\
||\bar{x}^T\bar{x}|| &= \sum_{i=1}^{n} \bar{x}_i^2 \\
&= \frac{\partial \bar{x}^2}{\partial f} \\
\nabla_{\bar{x}}(\bar{x}^T\bar{x}) &= 2\bar{x} \\
&\quad \text{Since c is scalar the gradient is 0, so the gradient is just} \\
\nabla_{\bar{x}}(\bar{x}^T\bar{x}) &= 2\bar{x}
\end{aligned}
$$

2.2 b

$$
\begin{aligned}
f(\bar{x}) &= \bar{x}^T\bar{b} \\
&\quad \text{Gradient of } \bar{x}^T\bar{b} \\
||\bar{x}^T\bar{b}|| &= \sum_{i=1}^{n} \bar{x}_i\bar{b}_i \\
&= \frac{\partial \bar{x}\bar{b}}{\partial f} \\
\nabla_{\bar{x}}(\bar{x}^T\bar{b}) &= \bar{b}
\end{aligned}
$$

2.3 c

$$\begin{aligned}
f(\bar{x}) &= \bar{x}^T A\bar{x} + \bar{b}^T\bar{x} + c \\
&\text{Gradient of } \bar{x}^T A\bar{x} \\
||\bar{x}^T A\bar{x}|| &= A\sum_{i=1}^{n} \bar{x}_i^2 \\
&= A \cdot \frac{\partial \bar{x}^2}{\partial f} \\
\nabla_{\bar{x}}(\bar{x}^T\bar{x}) &= A2\bar{x} \\
&\text{Gradient of } \bar{b}^T\bar{x} \\
||\bar{b}^T\bar{x}|| &= \sum_{i=1}^{n} \bar{x}_i\bar{b}_i \\
&= \frac{\partial \bar{x}\bar{b}}{\partial f} \\
\nabla_{\bar{x}}(\bar{x}^T\bar{b}) &= \bar{b} \\
&\text{Combine} \\
\nabla_{\bar{x}}(f\bar{x}) &= 2A\bar{x} + \bar{b}
\end{aligned}$$

3 Exercise 3

3.1 a

```
mean_compute = t_train.mean()
ptint(mean_compute)
-----output-----
22.016600790513834
```

3.2 b

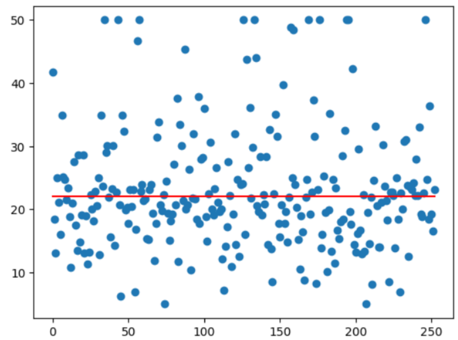
```
def rmse(y_true, y_pred):
    return numpy.sqrt(numpy.mean((y_true - y_pred) ** 2))

rmse_loss = rmse(t_test, mean_compute)
print(rmse_loss)
-----output-----
9.672477972746307
```

3.3 c

```
import matplotlib.pyplot as plt

plt.scatter(range(len(t_test)), t_test)
plt.plot(range(len(t_test)), [mean_compute] * len(t_test), color = "red")
plt.show()
```



4 Exercise 4

4.1 a)

```
ones = numpy.c_[numpy.ones(X.shape[0]), X]
self.w = numpy.linalg.pinv(ones.T @ ones) @ ones.T @ t
```

To find the w I used the numpy.c_ function that takes two np.arrays of the same length, transposes them, and creates a new matrix (2xlength of the arrays). This allows me to find the inverse of $X^T X$, which then allows me to find the w: $w = (X^T X)^{-1} X^T \cdot t$. These part I added to the fit function.

4.2 b)

```
model_single = linreg.LinearRegression()
model_single.fit(X_train[:,0], t_train)
print(model_single.w)
-----output-----
[[23.63506195]
 [-0.43279318]]
```

This indicates that the output is very dependent on the first feature, meaning to if this one increases, the output will increase significantly.

4.3 c)

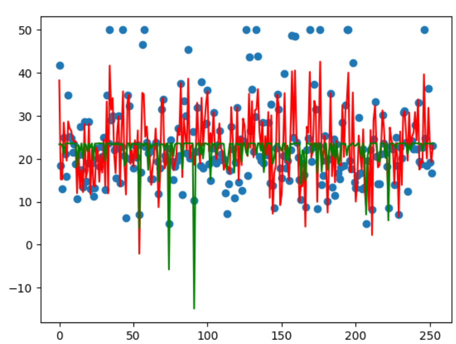
```
model = linreg.LinearRegression()
model.fit(X_train, t_train)
print(model.w)
-----output-----
[[ 3.13886978e+01]
 [-5.96169127e-02]
 [ 2.93672792e-02]
 [-2.90605834e-02]
 [ 2.29256181e+00]
 [-1.73263655e+01]
 [ 3.99375996e+00]
 [ 3.23077761e-03]
 [-1.28724508e+00]
 [ 3.54780191e-01]
 [-1.55819191e-02]
 [-8.14647713e-01]
 [ 1.17820208e-02]
 [-4.64869014e-01]]
```

4.4 d)

```
p = model.predict(X_test)
p_single = model_single.predict(X_test[:,0])

rmse_loss = rmse(t_test, p)
rmse_loss_single = rmse(t_test, p_single)
print(rmse_loss, rmse_loss_single)

plt.scatter(range(len(t_test)), t_test)
plt.plot(range(len(t_test)), p, color = "red")
plt.plot(range(len(t_test)), p_single, color = "green")
plt.show()
-----output-----
4.688333653614856 8.954859906611233
```



Here you can see the output graph where the green is the single prediction model and the red is using it all. As you can see the RMSE is half when using all the features then when it uses just one.

5 Exercise 5

Consider:

$$\mathcal{L} = \sum_{n=1}^{N} (f(x_n; w) - t_n)^2$$
$$\mathcal{L} = ||Xw - t||^2$$

Here:

- X the matrix of stacked inputs x_n of size Nxd
- w is a d-dimensional vector
- t is a N-dimensional vector

If we then take the partail derivative of w:

$$\frac{\partial \mathcal{L}}{\partial w} = (||Xw - t||^2)'$$
$$\frac{\partial \mathcal{L}}{\partial w} = 2X^T(Xw - t)$$

Find solution for when the derative is 0 (first-order optimality condition)

$$X^T(Xw - t) = 0$$
$$X^TXw - X^Tt = 0$$
$$X^TXw = X^Tt$$
$$w = (X^TX)^{-1}X^Tt$$

Which should then be the optimal least squared parameter vector. Comparing to the average loss, which is $\frac{1}{N}\mathcal{L}$, the only difference is the $\frac{1}{N}$, which shouldn't effect the optimization steps.