DSC 40A - Homework 2

Due: Friday, January 21, 2022 at 11:59pm

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Homeworks are due to Gradescope by 11:59pm on the due date. You can use a slip day to extend the deadline by 24 hours.

Homework will be evaluated not only on the correctness of your answers, but on your ability to present your ideas clearly and logically. You should **always explain and justify** your conclusions, using sound reasoning. Your goal should be to convince the reader of your assertions. If a question does not require explanation, it will be explicitly stated.

Homeworks should be written up and turned in by each student individually. You may talk to other students in the class about the problems and discuss solution strategies, but you should not share any written communication and you should not check answers with classmates. You can tell someone how to do a homework problem, but you cannot show them how to do it.

For each problem you submit, you should **cite your sources** by including a list of names of other students with whom you discussed the problem. Instructors do not need to be cited.

This homework will be graded out of 50 points. The point value of each problem or sub-problem is indicated by the number of avocados shown.

Note: For Problem 5, part (c), you'll need to use a supplementary notebook, linked here. We won't grade your code this time; you just need to submit some screenshots of your code's output and a plot, which you can submit along with the rest of the assignment.

Problem 1. Linear Transformations

Suppose we are given a data set $\{d_1, d_2, \ldots, d_n\}$ and know its mean, variance, and standard deviation to be $mean_d$, var_d, and std_d. Consider another data set $\{t_1, t_2, \ldots, t_n\}$, where t_i is a linear transformation of d_i:

$$t_i = f(d_i) = a \cdot d_i + b$$

for each $i = 1, 2, \ldots, n$. Here, a and b are arbitrary constants. Let $mean_t$, var_t, and std_t be the mean, variance, and standard deviation of the transformed data.

a) Express $mean_t$ in terms of $mean_d$, a, and b.

Solution:

$$\begin{aligned} mean_t &= \frac{1}{n}\sum_{i=1}^{n}(a \cdot d_i + b) \\ &= a \cdot \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right) + b \\ &= a \cdot mean_d + b \end{aligned}$$

b) Express var_t in terms of var_d, a, and b.

Solution:

$$\begin{aligned}
var_t &= \frac{1}{n}\sum_{i=1}^{n}(a \cdot d_i + b - mean_t)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(a \cdot d_i + b - (a \cdot mean_d + b))^2 \\
&= a^2 \cdot \frac{1}{n}\sum_{i=1}^{n}(d_i - mean_d)^2 \\
&= a^2 \cdot var_d
\end{aligned}$$

c) Express std_t in terms of std_d, a, and b.

Solution:

$$\begin{aligned}
std_t &= \sqrt{var_t} \\
&= \sqrt{a^2 \cdot var_d} \\
&= |a| \cdot \sqrt{var_d} \\
&= |a| \cdot std_d.
\end{aligned}$$

Problem 2. Quadratic Mean

Suppose we are given a data set of size n with $0 < y_1 \leq y_2 \leq \cdots \leq y_n$.

Define a new loss function by

$$L_Q(h, y) = (h^2 - y^2)^2$$

and consider the empirical risk

$$R_Q(h) = \frac{1}{n}\sum_{i=1}^{n} L_Q(h, y_i).$$

a) Show that $R(h)$ has critical points at $h = 0$ and when h equals the **quadratic mean** of the data, defined as

$$QM(y_1, y_2, \ldots, y_n) = \sqrt{\frac{y_1^2 + y_2^2 + \cdots + y_n^2}{n}}.$$

Solution: We start by finding the critical points of $R_Q(h)$.

$$\begin{aligned}
R'_Q(h) &= \frac{d}{dh}\left(\frac{1}{n}\sum_{i=1}^{n}(h^2 - y_i^2)^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{d}{dh}(h^2 - y_i^2)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}2(h^2 - y_i^2)\cdot 2h \\
&= \frac{4h}{n}\sum_{i=1}^{n}(h^2 - y_i^2) \\
&= \frac{4h}{n}\left(nh^2 - \sum_{i=1}^{n}y_i^2\right)
\end{aligned}$$

Let $R'_Q(h) = 0$, then either $\dfrac{4h}{n} = 0$, or $nh^2 - \displaystyle\sum_{i=1}^{n} y_i^2 = 0$.

Therefore, $R_Q(h)$ has critical points when:

$$h = 0 \text{ and } h = \sqrt{\frac{y_1^2 + y_2^2 + \cdots + y_n^2}{n}} = QM(y_1, y_2, \cdots, y_n).$$

b) Recall from single-variable calculus the **second derivative test**, which says that for a function f with critical point at x^*,

- if $f''(x^*) > 0$, then x^* is a local minimum, and
- if $f''(x^*) < 0$, then x^* is a local maximum.

Use the second derivative test to determine whether each critical point you found in part (a) is a maximum or minimum of $R_Q(h)$.

Solution: For a critical point h to be a minimum, we must show that $R''_Q(h) > 0$. For a critical point to be a maximum, we must show that $R''_Q(h) < 0$.

We start by rewriting the first derivative and taking the second derivative:

$$\begin{aligned}
R'_Q(h) &= \frac{4}{n}\left(nh^3 - h\sum_{i=1}^{n}y_i^2\right) \\
R''_Q(h) &= \frac{4}{n}\left(\frac{d}{dh}\left(nh^3\right) - \frac{d}{dh}\left(h\sum_{i=1}^{n}y_i^2\right)\right) \\
&= \frac{4}{n}\left(3nh^2 - \sum_{i=1}^{n}y_i^2\right) \\
&= 12h^2 - \frac{4}{n}\sum_{i=1}^{n}y_i^2
\end{aligned}$$

When $h = 0$, $R_Q''(h) = -\frac{4}{n}\sum_{i=1}^{n} y_i^2 < 0$. Therefore $h = 0$ is a local maximum.

When h is the quadratic mean, $h^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2$, so

$$\begin{aligned}
R_Q''(h) &= 12h^2 - \frac{4}{n}\sum_{i=1}^{n} y_i^2 \\
&= \frac{12}{n}\sum_{i=1}^{n} y_i^2 - \frac{4}{n}\sum_{i=0}^{n} y_i^2 \\
&= \frac{8}{n}\sum_{i=1}^{n} y_i^2 \\
&> 0 \text{ because } 0 < y_1 \leq y_2 \leq \cdots \leq y_n.
\end{aligned}$$

Therefore, $h = QM(y_1, y_2, \ldots, y_n)$ is a local minimum.

c) Show that the quadratic mean always falls between the smallest and largest data values, which is a property that any reasonable prediction should have. This amounts to proving the inequality

$$y_1 \leq QM(y_1, y_2, \ldots, y_n) \leq y_n.$$

Solution: We start by proving $y_1 \leq QM(y_1, y_2, \ldots, y_n)$.

Given $0 < y_1 \leq y_2 \leq \cdots \leq y_n$, then we know $y_1 \leq y_i$ for $i = 1, 2, \ldots, n$. This means $y_1^2 \leq y_i^2$ for $i = 1, 2, \ldots, n$.

Therefore,

$$\begin{aligned}
QM(y_1, y_2, \ldots, y_n) &= \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n}} \\
&\geq \sqrt{\frac{\sum_{i=1}^{n} y_1^2}{n}} \\
&\sqrt{\frac{n \cdot y_1^2}{n}} \\
&= \sqrt{y_1^2} \\
&= y_1
\end{aligned}$$

This shows $QM(y_1, y_2, \ldots, y_n) \geq y_1$.

The second part of the inequality can be proved similarly. We are given that $0 < y_1 \leq y_2 \leq \cdots \leq y_n$, which implies $y_n \geq y_i$ for $i = 1, 2, \ldots, n$. This means $y_n^2 \geq y_i^2$ for $i = 1, 2, \ldots, n$.

Therefore,

$$\begin{aligned} QM(y_1, y_2, \ldots, y_n) &= \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n}} \\ &\le \sqrt{\frac{\sum_{i=1}^{n} y_n^2}{n}} \\ &\sqrt{\frac{n \cdot y_n^2}{n}} \\ &= \sqrt{y_n^2} \\ &= y_n \end{aligned}$$

This shows $QM(y_1, y_2, \ldots, y_n) \le y_n$.

Problem 3. Happy Family

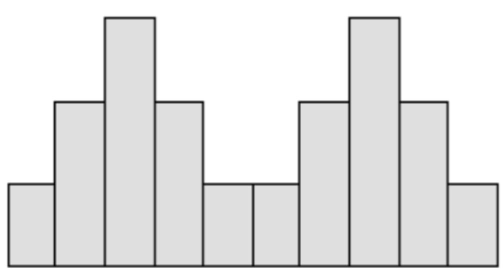
In class, we defined the *mean absolute deviation from the median* as a measure of the spread of a data set. This measure takes the absolute deviations, or differences, of each value in the data set from the median, and computes the mean of these absolute deviations. We can think of this one measure of spread as a member of a family of analogously defined measures of spread:

- mean absolute deviation from the median
- median absolute deviation from the median
- mean absolute deviation from the mean
- median absolute deviation from the mean

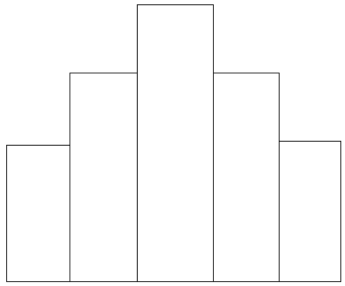
While all four of these measures capture the notion of spread, they do so in different ways, and so they may have different values for the same data set.

a) For the data set whose histogram is shown below, draw a histogram showing the rough shape of the distribution of the absolute deviations from the mean. Which of these two measures is greater, or are they about the same?

- mean absolute deviation from the mean
- median absolute deviation from the mean



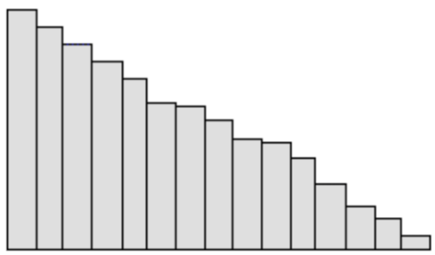
Solution: The mean is in the center of the histogram since it is roughly symmetric. The histogram of absolute deviations from the mean is



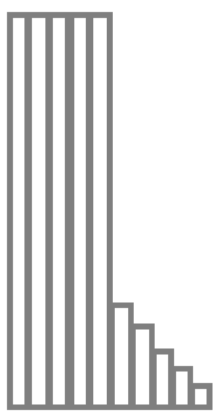
Since the histogram of absolute deviations from the mean is roughly symmetric, we know its mean value is about the same as its median. Therefore, the mean absolute deviation from the mean should be about the same as the median absolute deviation from the mean.

b) For the data set whose histogram is shown below, draw a histogram showing the rough shape of the distribution of the absolute deviations from the median. Which of these two measures is greater, or are they about the same?

- mean absolute deviation from the median
- median absolute deviation from the median



Solution: The median of the histogram is slightly left of center, since the histogram is right-skewed. We can draw the histogram of absolute deviations from the median, which gives



Since the histogram of absolute deviations from the median is right-skewed, we know its mean value is greater than its median. Therefore, the mean absolute deviation from the median should be greater than the median absolute deviation from the median.

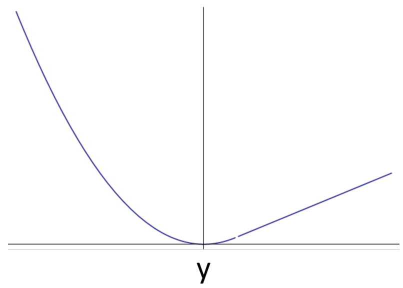
Problem 4. Piecewise Loss

Consider a new loss function,

$$L_p(h, y) = \begin{cases} (h-y)^2, & h \leq y + \frac{1}{2} \\ h - (y + \frac{1}{4}), & h > y + \frac{1}{2} \end{cases}.$$

a) Fix an arbitrary value of y. Draw the graph of $L_p(h, y)$ as a function of h. You should notice that $L_p(h, y)$ is minimized at y.

Solution:



b) Recall from single-variable calculus the definition of **continuity at a point**:

$$f(x) \text{ is continuous at } x = a \text{ if } \lim_{x \to a} f(x) \text{ exists and is equal to } f(a).$$

Also, remember that for $\lim_{x \to a} f(x)$ to exist, the left-hand limit and right-hand limit must match:

$$\lim_{x \to a^-} f(x) = \lim_{x \to a^+} f(x).$$

Fix an arbitrary value of y. Show that as a function of h, $L_p(h, y)$ is continuous for all h.

Solution: Since each piece of $L_p(h, y)$ is a continuous function of h, we only need to make sure that $L_p(h)$ is continuous at $h = y + \frac{1}{2}$ by showing that the left-hand limit and right-hand limit are the same, and that the function value at that point is also the same.

We have that

$$\begin{aligned} \lim_{h \to (y + \frac{1}{2})^-} L_p(h, y) &= \lim_{h \to (y + \frac{1}{2})^-} (h - y)^2 \\ &= \left(y + \frac{1}{2} - y\right)^2 \\ &= \frac{1}{4} \end{aligned}$$

using the first branch of the function.

Similarly,

$$\begin{aligned}
\lim_{h \to (y+\frac{1}{2})^+} L_p(h,y) &= \lim_{h \to (y+\frac{1}{2})^+} h - \left(y + \frac{1}{4}\right) \\
&= y + \frac{1}{2} - \left(y + \frac{1}{4}\right) \\
&= \frac{1}{4}
\end{aligned}$$

using the second branch of the function.

Since the left and right limits are the same, $\lim_{h \to (y+\frac{1}{2})} L_p(h,y)$ exists and is equal to $\frac{1}{4}$, and this is the same as $f(y+\frac{1}{2}) = (y + \frac{1}{2} - y)^2 = \frac{1}{4}$.

So $L_p(h,y)$ is continuous at $h = y + \frac{1}{2}$ and therefore continuous for all h.

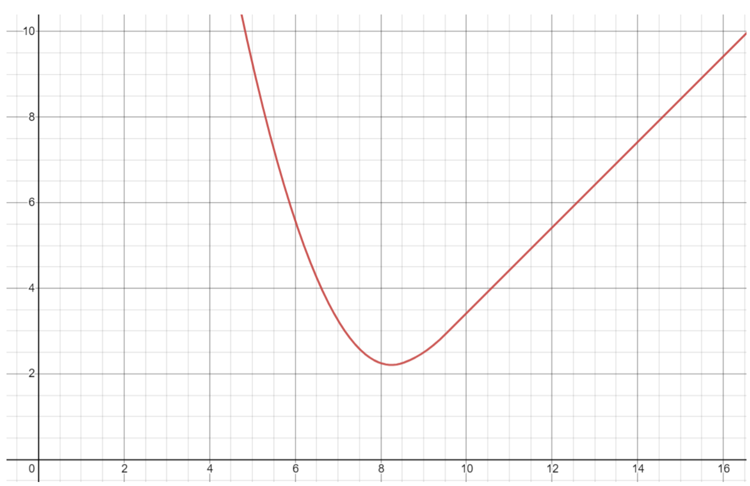
c) Again fix an arbitrary value of y. The function $L_p(h,y)$ is differentiable for all h. Calculate its derivative $L_p'(h,y)$, which will be a piecewise function of h, and show that both pieces of the function evaluate to the same value at the transition point $h = y + \frac{1}{2}$.

Solution: We can take the derivative with respect to h of each piece to obtain that

$$L_p'(h,y) = \begin{cases} 2(h-y), & h \le y + \frac{1}{2} \\ 1, & h > y + \frac{1}{2} \end{cases}.$$

Note that when $h = y + \frac{1}{2}$, we have $2(h-y) = 2(y + \frac{1}{2} - y) = 1$, so both portions of the function evaluate to the same value.

d) Suppose our data set is $\{2, 8, 9\}$. The plot of the empirical risk for this dataset, $R_p(h) = \frac{1}{n}\sum_{i-1}^{n} L_p(h, y_i)$, is shown below:



It is not possible to directly solve for the value of h which minimizes this function. Instead, run gradient descent by hand using an initial prediction of $h_0 = 9$ and a step size of $\alpha = \frac{3}{4}$. Run the algorithm until it converges (it shouldn't take too many iterations). Please show your calculations, and to help the graders track your progress, include a boxed summary with the value of h at each iteration, such as below:

$$
\begin{array}{|l|}
\hline
h_0 = 9 \\
h_1 = \dots \\
h_2 = \dots \\
\vdots \\
\hline
\end{array}
$$

Solution: First we must calculate the derivative of the risk. We have:

$$
\begin{aligned}
\frac{dR_\text{p}}{dh}(h) &= \frac{d}{dh}\left[\frac{1}{n}\sum_{i=1}^{n} L_\text{p}(h, y_i)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n} \frac{dL_\text{p}}{dh}(h, y_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} \begin{cases} 2(h-y), & h \le y + \frac{1}{2} \\ 1, & h > y + \frac{1}{2} \end{cases}
\end{aligned}
$$

We start the first iteration with $h_0 = 9$. To apply the gradient descent update rule, we first have to calculate the derivative of R_p at $h_0 = 9$:

$$
\begin{aligned}
\frac{dR_\text{p}}{dh}(9) &= \frac{1}{3}\sum_{i=1}^{n} \frac{dL_\text{p}}{dh}(9, y_i) \\
&= \frac{1}{3}\left[\frac{dL_\text{p}}{dh}(9,2) + \frac{dL_\text{p}}{dh}(9,8) + \frac{dL_\text{p}}{dh}(9,9)\right]
\end{aligned}
$$

We now evaluate each derivative using the answer from the previous part.

$$
\begin{aligned}
&= \frac{1}{3}\left[1 + 1 + 2\cdot(9-9)\right] \\
&= \frac{2}{3}
\end{aligned}
$$

Applying the update rule, we find:

$$
\begin{aligned}
h_1 &= h_0 - \alpha\frac{dR_\text{p}}{dh}(h_0) \\
&= 9 - \frac{3}{4}\cdot frac23 \\
&= 8.5
\end{aligned}
$$

On to the second iteration. We start by calculating the slope at $h_1 = 8.5$:

$$
\begin{aligned}
\frac{dR_\text{p}}{dh}(h_1) &= \frac{dR_\text{p}}{dh}(8.5) \\
&= \frac{1}{3}\sum_{i=1}^{n} \frac{dL_\text{p}}{dh}(8.5, y_i) \\
&= \frac{1}{3}\left[\frac{dL_\text{p}}{dh}(8.5, 2) + \frac{dL_\text{p}}{dh}(8.5, 8) + \frac{dL_\text{p}}{dh}(8.5, 9)\right] \\
&= \frac{1}{3}\left[1 + 2 \cdot (8.5 - 8) + 2 \cdot (8.5 - 9)\right] \\
&= \frac{1}{3}
\end{aligned}
$$

Applying the update rule, we find:

$$
\begin{aligned}
h_2 &= h_1 - \alpha \frac{dR_\text{p}}{dh}(h_1) \\
&= 8.5 - \frac{3}{4} \cdot \frac{1}{3} \\
&= 8.25
\end{aligned}
$$

On the third iteration, we have:

$$
\begin{aligned}
\frac{dR_\text{p}}{dh}(h_2) &= \frac{dR_\text{p}}{dh}(8.25) \\
&= \frac{1}{3}\sum_{i=1}^{n} \frac{dL_\text{p}}{dh}(8.25, y_i) \\
&= \frac{1}{3}\left[\frac{dL_\text{p}}{dh}(8.25, 2) + \frac{dL_\text{p}}{dh}(8.25, 8) + \frac{dL_\text{p}}{dh}(8.25, 9)\right] \\
&= \frac{1}{3}\left[1 + 2 \cdot (8.25 - 8) + 2 \cdot (8.25 - 9)\right] \\
&= \frac{1}{3}\left[1 + 0.5 + -1.5\right] \\
&= 0
\end{aligned}
$$

Applying the update rule, we find:

$$
\begin{aligned}
h_3 &= h_2 - \alpha \frac{dR_\text{p}}{dh}(h_2) \\
&= 8.25 - \frac{3}{4} \cdot 0 \\
&= 8.25
\end{aligned}
$$

We have converged to 8.25 after three iterations. The progress of the algorithm was:

$$
\boxed{\begin{aligned}
h_0 &= 9 \\
h_1 &= 8.5 \\
h_2 &= 8.25 \\
h_3 &= 8.25
\end{aligned}}
$$

e) Does the prediction for the dataset $\{2, 8, 9\}$ seem high or low to you? Why do you think this is the case, based on the graph of the piecewise loss function you drew in part (a)?

Solution: A prediction of 8.25 is pretty high for the dataset $\{2, 8, 9\}$. It's higher than both the mean and the median, for example. This happens because $L_p(h, y)$ is much larger when the prediction h is less than the data value y, as compared to when the prediction is larger than the data value. The graph of $L_p(h, y)$ is not symmetric, and is steeper on the left of y than on the right. This means the loss is greater for lower predictions, and so predictions made with this loss function tend to be on the high side.

Problem 5. Gradient Descent, Linear Regression, and Recipe Nutrition

Soon in the course we are going to learn about linear regression, which is commonly known as finding the "line of best fit". Interestingly enough, the metric used to define "best fit" is a function you are probably quite familiar with already, the squared risk function, also called the mean squared error:

$$R_{sq}(h) = \frac{1}{n}\sum_{i=1}^{n}(y_i - h)^2.$$

So far in this class, our prediction h has been a single real number. When performing linear regression, however, our prediction is allowed to vary with each x_i, according to some linear function $f(x_i) = mx_i + b$, where m and b are real numbers. That means, when talking about how well a linear function "fits" the data, we are measuring the fit by the mean squared error:

$$R_{sq}(f) = R_{sq}(m, b) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (mx_i + b))^2$$

and linear regression is just finding the values of m and b that minimize this mean squared error.

While this may look intimidating, we actually have already learned an excellent tool that will be able to help us solve this problem: gradient descent.

Gradient descent is guaranteed to find the global minimum of a function if the function is *convex* (also called concave up) and *differentiable*. In order to verify that gradient descent will indeed find the global minimum for us, we need to prove that the mean squared error is both convex and differentiable. Remember from calculus that a function $g : \mathbb{R} \to \mathbb{R}$ is convex (concave up) if

$$g''(x) = \frac{\mathrm{d}^2 g}{\mathrm{d}x^2} \geq 0, \text{for all } x.$$

a) Use the definition of convexity above to show that

$$R_{sq}(h) = \frac{1}{n}\sum_{i=1}^{n}(y_i - h)^2$$

is a convex function of h.

Solution: We compute the second derivative of $R_{sq}(h)$ and check to make sure that it is positive no matter the value of h:

$$\begin{aligned}
\frac{\mathrm{d}^2 R_{sq}}{\mathrm{d}h^2} &= \frac{\mathrm{d}^2}{\mathrm{d}h^2}\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - h)^2\right) \\
&= \frac{\mathrm{d}^2}{\mathrm{d}h^2}\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2 - 2y_i h + h^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{d}^2}{\mathrm{d}h^2}\left(y_i^2 - 2y_i h + h^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} 2 \\
&= \frac{1}{n}\cdot 2n \\
&= 2.
\end{aligned}$$

We have that $2 \geq 0$ for all values of h, so $R_{sq}(h)$ is convex.

b) In order to run gradient descent on $R_{sq}(m, b)$, we first need to find it's gradient. Thinking of $R_{sq}(m, b)$ as a function of two variables, compute the partial derivatives $\frac{\partial}{\partial m} R_{sq}(m, b)$ and $\frac{\partial}{\partial b} R_{sq}(m, b)$.

Solution:

Recall that the expression for $R_{sq}(m, b)$ is

$$R_{sq}(m, b) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (mx_i + b))^2$$

which we can expand as

$$= \frac{1}{n}\sum_{i=1}^{n}\left(b^2 + 2bmx_i - 2by_i + m^2x_i^2 - 2mx_iy_i + y_i^2\right).$$

This expanded version will be useful to simplify the partial derivatives. Alternatively, without expanding, we could also take derivatives by applying the chain rule.

We compute the partial derivatives as follows:

$$\begin{aligned}
\frac{\partial}{\partial m} R_{sq}(m,b) &= \frac{\partial}{\partial m}\left(\frac{1}{n}\sum_{i=1}^{n}\left(b^2 + 2bmx_i - 2by_i + m^2x_i^2 - 2mx_iy_i + y_i^2\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial m}\left(b^2 + 2bmx_i - 2by_i + m^2x_i^2 - 2mx_iy_i + y_i^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(2bx_i + 2mx_i^2 - 2x_iy_i\right) \\
&= \frac{2}{n}\sum_{i=1}^{n} -x_i(y_i - (mx_i + b)) \\
\frac{\partial}{\partial b} R_{sq}(m,b) &= \frac{\partial}{\partial b}\left(\frac{1}{n}\sum_{i=1}^{n}\left(b^2 + 2bmx_i - 2by_i + m^2x_i^2 - 2mx_iy_i + y_i^2\right)\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial b}\left(b^2 + 2bmx_i - 2by_i + m^2x_i^2 - 2mx_iy_i + y_i^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(2b + 2mx_i - 2y_i\right) \\
&= \frac{2}{n}\sum_{i=1}^{n} -(y_i - (mx_i + b)).
\end{aligned}$$

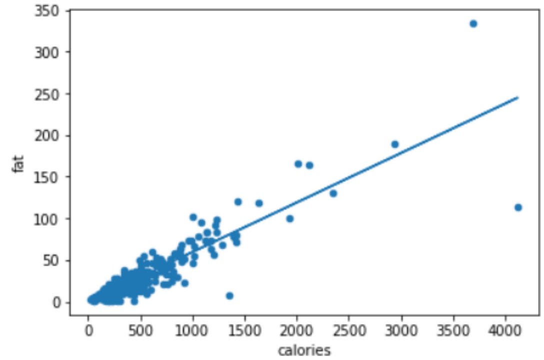
c) Now, let's try using all this math to find a regression line for some real data. We'll be looking at nutritional information from recipes on Epicurious and trying to predict the amount of fat based on the number of calories.

In this supplementary notebook (linked), fill in the missing functions based on your answer to part (b), read through the code we've provided, and follow the instructions to submit screenshots of your work. We won't grade your code this time; you just need to submit certain screenshots, which you can submit along with the rest of the assignment.

Solution: The output is shown here:

```
Performing linear regression with gradient descent...
After  1000  iterations, m = 0.05942135354003322 and b =  -0.0006066747975734924
After  2000  iterations, m = 0.059422044331360925 and b =  -0.0012745471855237962
After  3000  iterations, m = 0.05942273485162615 and b =  -0.001942157504252416
After  4000  iterations, m = 0.059423425100935255 and b =  -0.0026095058565937844
After  5000  iterations, m = 0.059424115079394574 and b =  -0.0032765923453419873
After  6000  iterations, m = 0.059424804787110375 and b =  -0.003943417073250763
After  7000  iterations, m = 0.05942549422418889 and b =  -0.004609980143033595
After  8000  iterations, m = 0.05942618339073633 and b =  -0.005276281657363432
After  9000  iterations, m = 0.059426872286858855 and b =  -0.005942321718873261
After  10000  iterations, m = 0.05942756091266255 and b =  -0.006608100430155601
Gradient descent completed.
The slope is m = 0.05942756091266255
The y-intercept is b = -0.006608100430155601
```

The regression line and data are plotted here:



Below is the code that you needed to fill in:

```
def m_gradient_func(x_i, y_i, m, b, n):
    '''Return the single term in the gradient of m corresponding to (x_i, y_i).'''
    return -(2 / n) * x_i * (y_i - ((m * x_i) + b))
```

```
def b_gradient_func(x_i, y_i, m, b, n):
    '''Return the single term in the gradient of b corresponding to (x_i, y_i).'''
    return -(2 / n) * (y_i - ((m * x_i) + b))
```