Notebook

April 4, 2020

```
In [3]: def mean(population):
            """
            Returns the mean of population (mu)

            Keyword arguments:
            population -- a numpy array of numbers
            """
            # Calculate the mean of a population
            ...

        def variance(population):
            """
            Returns the variance of population (sigma squared)

            Keyword arguments:
            population -- a numpy array of numbers
            """
            # Calculate the variance of a population
            ...
```

```
In [9]: v = np.array([2,1,4])

        B = ...

        # The notation B @ v means: compute the matrix multiplication Bv
        B @ v


        ---------------------------------------------------------------------------
```

```
In [ ]: A = np.array([
            [2, 1, 0, 0],
            # Finish this!
            ...
        ])

        A @ B @ v
```

```
In [ ]: most = ...
```

```
In [ ]: new_v = ...
        new_v
```

0.1 Question 3

For each of the statements below, either prove that it is true by using definitions or show that it is false by providing a counterexample.

Statement I $\frac{\sum_{i=1}^{n} a_i x_i}{\sum_{i=1}^{n} a_i} = \sum_{i=1}^{n} x_i$

Write your answer here, replacing this text.

Statement II $\sum_{i=1}^{n} x_1 = nx_1$

Write your answer here, replacing this text.

Statement III $\sum_{i=1}^{n} a_3 x_i = n a_3 \bar{x}$

Write your answer here, replacing this text.

Statement IV $\sum_{i=1}^{n} a_i x_i = n\bar{a}\bar{x}$

Write your answer here, replacing this text.

1 Gradients and Differentiation

1.1 Question 4a

Suppose we have the following scalar-valued function on x and y:

$$f(x, y) = 2x^2 + 3xy + y^3 + e^{-2y} + \ln(4y)$$

Compute the partial derivative of $f(x, y)$ with respect to x.
Write your answer here, replacing this text.

Now compute the partial derivative of $f(x, y)$ with respect to y:
Write your answer here, replacing this text.

Finally, using your answers to the above two parts, compute $\nabla f(x, y)$ (the gradient of $f(x, y)$) and evaluate the gradient at the point $(x = 2, y = -1)$.

Write your answer here, replacing this text.

1.2 Question 4b

Use calculus to find the value(s) of θ which minimizes the expression below. Justify why it is the minimum.

$\sum_{x=1}^{10}(x-\theta)^2$

Write your answer here, replacing this text.

```
In [ ]: ## Fill in the functions below
        def f(x):
            ...

        def df0(x):
            ...

        def df8(x):
            ...

        ## Plot the result of the input function g
        ## Optionally set the color with the second argument
        ## Fill in x = ...
        def plot_function(g, col="blue"):

            # Set x-values to integer values of from -15 to 15
            x = ...

            source = pd.DataFrame({
              'x': x,
              'g(x)': g(x),
            })

            return alt.Chart(source).mark_line(color=col, clip=True).encode(
                x=alt.X('x',
                    scale=alt.Scale(domain=(-15, 15))
                ),
                y=alt.Y('g(x)',
                    scale=alt.Scale(domain=(-100, 300))
                )
            ).properties(
            width=600,
            height=300
        )

        p1 = plot_function(f, col="red")
        p2 = plot_function(df0)
        p3 = plot_function(df8, col="green")

        ## Combine all lines into one plot
        p1 + p2 + p3
```

2 Baye's Rule

2.1 Question 5

Consider the following scenario:

Only 1% of 40-year-old women who participate in a routine mammography test have breast cancer. 80% of women who have breast cancer will test positive, but 9.6% of women who don't have breast cancer will also get positive tests.

Suppose we know that a woman of this age tested positive in a routine screening. What is the probability that she actually has breast cancer?

You **must** show work using LaTex (not code) to get credit for your answer.

Hint: Use Bayes' rule.

Write your answer here, replacing this text.

3 Reading Documentation

3.0.1 Question 6

We should also familiarize ourselves with looking up documentation and learning how to read it. Below is a section of code that plots a basic wireframe. Replace each `# Your answer here` with a description, **in your own words** of what the line above does, what the arguments being passed in are, and how the arguments are used in the function. For example,

```
np.arange(2, 5, 0.2)
# This returns an array of numbers from 2 to 5 with an interval size of 0.2
```

Hint: The `Shift + Tab` tip from earlier in the notebook may help here. Remember that objects must be defined in order for the documentation shortcut to work; for example, all of the documentation will show for method calls from `np` since we've already executed `import numpy as np`. You may also want to experiment in the scratch cell to evaluate `x` and `xvec` (and similary for `y` and `z`) to see what the code is doing.

```
In [ ]: x, y = np.meshgrid(range(-5, 5), range(-5, 5))
        # Your answer here

        z = x ** 2 + y ** 2

        xvec = x.ravel()
        yvec = y.ravel()
        zvec = z.ravel()
        # You answer here


        ## The following code uses altair to make a heatmap
        ## We'll be exploring this more in future lectures / labs
        ## In this case altair requires 3 columns: an x coordinate, a y coordinate, and a third column
        source = pd.DataFrame({'x': xvec,
                               'y': yvec,
                               'z': zvec})

        alt.Chart(source).mark_rect().encode(
            x='x:O',
            y='y:O',
            color='z:Q'
        )
```