Part I: Data and Programming Ideas

1. Long data is "long" - typically there will be three columns, id, variable, and value, which would look something like this:

ID	Variable	Value
Barry Bonds	HR	762
Pete Rose	H	4256
Barry Bonds	RBI	1996

   Wide format data will have a row for each id, with the variables as columns. Pandas uses pivot to go from long to wide, and melt to go from wide to long.

2. Hint: series are the building blocks of dataframes

3. We talked about this in class on Tuesday 4/13.

4. Groupbys find all the pairs of matching values in a set of columns, and then perform aggregations on them. Can you think of a few examples?

5. Hint: make two series in a jupter notebook with partially overlapping indexes, and then add them together.

Part II: Probability

1. We talked about this in class on Tuesday 4/13.

2. There are two (equivalent) definitions: one involves conditional probability, and the other one involves the probability of an intersection.

3. We talked about this in class on Tuesday 4/13. Think about the venn diagram and what is being double counted!

4. We talked about this in class on Tuesday 4/13. How do I get Bayes theorem from the definition of conditional probability?

5. We can represent each outcome as pairs (i, j) where is the roll of the first

dice, j is the roll of the second one. Each pairs has probability (1/36)

$$
\begin{aligned}
P(X=5) &= P(\{(1,4)\cup(2,3)\cup(3,2)\cup(4,1)\}) \\
&= 1/36+1/36+1/36+1/36 \\
&= 1/9 \\
P(X_1=2|X=6) &= \frac{P(X_1=2\cap X=6)}{P(X=6)} \\
&= \frac{P((2,6))}{P(\{(1,5)\cup(2,4)\cup...\cup(5,1)\})} \\
&= \frac{1/36}{5/26} = 1/5 \\
P(X=6\cup X_2=3) &= P(X=6)+P(X_2=3)-P(X=6\cap X_2=3) \\
&= 5/36+1/6-1/36 \\
&= 5/18
\end{aligned}
$$

6. We talked about this in class on Tuesday 4/13. Countable and finite vs uncountable/continuum. Pmfs versus densities?

7. First, we find c so that this is a proper density:

$$
\begin{aligned}
\int_{-1}^{1} x^2 dx &= \frac{1}{3}x^3 \mid_{x=-1}^{x=1} \\
&= \frac{2}{3}
\end{aligned}
$$

so $c = 3/2$.

$$
\begin{aligned}
E[X] &= \frac{3}{2}\int_{-1}^{1} x^3 dx \\
&= \frac{3}{8}x^4 \mid_{x=-1}^{x=1} \\
&= 0 \\
E[X^2] &= \frac{3}{2}\int_{-1}^{1} x^4 dx \\
&= 3/5 \\
Var(X) &= E[X^2]-E[X]^2 \\
&= 3/5
\end{aligned}
$$

8. The expectation and variance can be calculated as follows:

$$\begin{aligned}
E[X] &= 0 \cdot 0.25 + 1 \cdot 0.5 + 2 \cdot 0.25 \\
&= 1 \\
E[X^2] &= 0^2 \cdot 0.25 + 1^2 \cdot 0.5 + 2^2 \cdot 0.25 \\
&= 1.5 \\
Var(X) &= 1.5 - 1^2 = 0.5
\end{aligned}$$

9. First we calculate the marginals:

$$\begin{aligned}
f_x(x) &= \int_0^1 (x+y)dy \\
&= x + \frac{1}{2} \\
f_y(y) &= \int_0^1 (x+y)dx \\
&= y + 1/2
\end{aligned}$$

for $x, y \in [0,1]$, 0 otherwise. So:

$$\begin{aligned}
E[X] &= \int_0^1 x(x+1/2)dx \\
&= 7/12 \\
E[X^2] &= \int_0^1 x^2(x+0.5)dx \\
&= 5/12 \\
Var(X) &= 5/12 - (7/12)^2 \\
&= 11/144 \\
SD(X) &= \sqrt{11/144}
\end{aligned}$$

The $E[Y]$,$Var(Y)$,and $SD(Y)$ will be the same. Next, we calculate:

$$\begin{aligned}
E[XY] &= \int_0^1 \int_0^1 xy * (x+y)dxdy \\
&= 1/3
\end{aligned}$$

So this give us:

$$
\begin{aligned}
Cov(X,Y) &= E[XY] - E[X]E[Y] \\
&= 1/3 - (7/12)^2 \\
&= -1/144 \\
Corr(X,Y) &= \frac{Cov(X,Y)}{SD(X)SD(Y)} \\
&= \frac{-1/144}{11/144} = -\frac{1}{11}
\end{aligned}
$$

Part III: Statistics

1. The likelihood and log-likelihood here is:

$$
\begin{aligned}
L(a|X_1,...,X_n) &= a^n \exp(-\frac{a}{2}\sum_{i=1}^{n} x_i^2)\prod_{i=1}^{n} x_i \\
\ell(a|X_1,...,X_n) &= n\log(a) - \frac{a}{2}\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n}\log(x_i)
\end{aligned}
$$

Taking the derivative with respect to a...

$$
\frac{\partial \ell}{\partial a} = \frac{n}{a} - \frac{1}{2}\sum_{i=1}^{n} x_i^2
$$

If we set this equal to zero and solve for a, we get:

$$
\hat{a} = \frac{2n}{\sum x_i^2}
$$

2. We will discuss on Thursday 4/15.

3. We will discuss on Thursday 4/15

4. The confidence interval here is:

$$
\begin{aligned}
(\bar{X} \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}) &= (1 \pm 2.262 \cdot \frac{\sqrt{0.8}}{\sqrt{10}}) \\
&\approx (.26, 1.74)
\end{aligned}
$$

We can reject H_0 because it is not in the confidence interval.

5. We will discuss on Thursday 4/15. But we've done this many times. Try to implement it yourself without copying and pasting code.

Part IV: Machine Learning

1. Think about the penalties that are being applied, and think about whether the coefficients are sparse or not.
2. Is this a regression or a classification problem. What is the formula for $p_i|\vec{X}_i$?
3. Be able to talk given k folds, what we do, and given a score, how we choose optimal hyperparameters.
4. Basically, be able to talk about how split the tree at a node, choosing the split that maximizes what score we're using. Understand what the main hyperparameters are. When do we stop splitting?
5. This should be pretty straightfoward...
6. This is a combination of bagging, and something else (the random part) when we grow the trees on each bootstrap sample
7. Well, for regression it is probably R^2, and for classification... well, it was "the hill I was willing to die on."
8. Practice doing this.