# 10-601 Machine Learning: Assignment 3

- The assignment is due at 3:00pm (beginning of class) on **Wednesday, Feb 20**.
- Write your name at the top right-hand corner of each page submitted.
- Each student must hand in a writeup. See the course webpage for the collaboration policy.
- For the programming portions of the assignment, you can use whatever programming language you are comfortable with. Submit code into the /hw3/ folder in your afs directory.
-

## 1 Q1 Logistic Regresion... [20 pts]

Exercise 3 from Naive Bayes and Logistic Regression reading.

## 2 Q2 ...And Gradient Descent [35 pts]

1. Download the portion of Fisher's Iris Flower data available on the class webpage. Implement logistic regression using gradient descent for weight updates, as discussed in the reading. Report the weights. Use 10-folds cross validation and report the classification accuracy.
2. Now repeat the experiment, this time using regularization as described in the reading. Vary your parameters $\eta$ and $\lambda$. What happens when you change these parameters?
3. Do you notice any differences in the results of the two different methods? Why or why not might that happen?

## 3 Q3 Statistical Tests [20 pts]

1. Depending on what we are trying to measure, we may choose to use a different hypothesis test.

   Suppose you had a set of webpages $\mathbf{X}$, labeled with the type of page $\mathbf{Y}$ (product, shopping cart, etc) $\mathbf{D} = D_{ij} =< x_{ij}, y_{ij} >$, representing sets of pages $i \in P$ from websites $j \in S$. You would like to compare hypotheses (in this case, classifiers) $h_A : X \rightarrow Y$ and $h_B : X \rightarrow Y$. You have two options for testing the hypotheses:

   Option 1: Split $\mathbf{D}$ into sets $D_{*1}, ..., D_{*S}$ where each set is all the pages from one website (note that some websites may have more pages than others). Do a paired test.

   Option 2: Split $\mathbf{D}$ evenly into $|S|$ different sets (each set containing pages sampled from all websites), and perform a paired test.

Option 3: Apply a test of $h_A$ and $h_B$ to all data in $\mathbf{D}$, and compare.

Which option would you expect to yield more accurate results for testing whether $h_A$ is more accurate than $h_B$ for future websites? Which option would be best for getting the true accuracy of $h_A$ and $h_B$? Explain your decisions.

2. Going back to the 20-sided die example, suppose that your friend's die scored a "critical hit" (19 or 20), 30 times out of 120 rolls. If $p_c$ is the probability of a critical hit ($p_c = 0.1$ in a fair die), what is the p-value of having a fair die in that situation? What if there were 50 critical hits in 200 rolls? Are the p-values the same? Why or why not?

3. Given a fair 20-sided die ($p = .1$), what is a 95% confidence interval for the number of critical hits you could expect out of 200 rolls? Out of 1000 rolls?