Quiz

SDS 385

1. Is independence of two random variables the same as uncorrelated? If not, what is the difference? No. Uncorrelated means $E[XY] = E[X]E[Y]$, whereas independent for two discrete R.V.s mean $P(X = x, Y = y) = P(X = x)P(Y = y)$. Independence implies uncorrelated but not vice versa. An example of uncorrelated but dependent RV's are as follows. For example, if $X \sim N(0, 1)$ and $Y \in \{-1, 1\}\}$ with probability $1/2$, then X, YX are uncorrelated, but they are not independent. You can check that they are uncorrelated by computing the expectation, but for checking independence, just note that $|X| = |YX|$, so they cannot be independent.

2. What is the difference between clustering and classification? The first has no labels provided, the second has labels.

3. How does the K-nearest neighbor algorithm for classification work? We did this in class.

4. Write down the definition of the largest singular value of a matrix as an optimization problem. For a matrix M, the largest singular value is given by $\sqrt{\max_{\|u\|=1} u^T M M^T u}$

5. Are the eigenvalues and singular values of a symmetric square matrix the same? If not, what is their relationship? Not necessarily, since eigenvalues can be negative but singular values are not. Singular values are absolute values of eigenvalues of a symmetric square matrix.

6. If f is convex, and X is a random variable, then how are $f(E[X])$ and $E[f(X)]$ related? Jensen's inequality.

7. Write down the definition of a symmetric positive-semi definite matrix. What do you know about its eigenvalues? For all $x \in \Re^n$ $x^T M x \geq 0$, where M is the $n \times n$ matrix in question. The eigenvalues are always positive.