ML course, 2016 fall
What you should know:

Week 1, 2 and 3.$\frac{1}{2}$: Basic issues in Probabilities and Information theory

Read: Chapter 2 from the *Foundations of Statistical Natural Language Processing* book by Christopher Manning and Hinrich Schütze, MIT Press, 2002, and/or *Probability Theory Review for Machine Learning*, Samuel Ieong, November 6, 2006 (https://see.stanford.edu/materials/aimlcs229/cs229-prob.pdf).

Week 1: Random events

(slides #3-#6 from https://profs.info.uaic.ro/∼ciortuz/SLIDES/foundations.pdf)

Concepts/definitions:

- event/sample space, random event
- probability function
- conditional probabilities
- independent random events (2 forms); conditionally independent random events (2 forms)

Theoretical results/formulas:

- elementary probability formula: $\frac{\text{\# favorable cases}}{\text{\# all possible cases}}$
- the “multiplication” rule; the “chain” rule
- “total probability” formula (2 forms)
- Bayes formula (2 forms)

Exercises illustrating the above concepts/definitions and theoretical results/formulas, in particular: proofs for certain properties derived from the *definition of the probability function* for instance: $P(\emptyset) = 0$, $P(\bar{A}) = 1 - P(A)$, $A \subseteq B \Rightarrow P(A) \leq P(B)$

Ciortuz et al.’s exercise book: ch. 1, ex. 1-5 [6-7], 8, 39-42 [43-45]

Week 2: Random variables

(slides #7-#16 [#35-#44] from https://profs.info.uaic.ro/∼ciortuz/SLIDES/foundations.pdf)

Concepts/definitions:

- random variables;
random variables obtained through function composition
- discrete random variables;
probability mass function (pmf)
examples: Bernoulli, binomial, geometric, Poisson distributions
- cumulative function distribution
- continuous random variables;
probability density function (pdf)
examples: Gaussian, exponential, Gamma distributions
- expectation (mean), variance, standard variation; covariance. (**See definitions!**)
- multi-valued random functions;
joint, marginal, conditional distributions
- independence of random variables;
conditional independence of random variables

Advanced issues:

- the likelihood function (see also Week 12)
- vector of random variables;
covariance matrix for a vector of random variables;
pozitive [semi-]definite matrices,
negative [semi-]definite matrices

Theoretical results/formulas:

- for any discrete variable X:
$\sum_x p(x) = 1$, where p is the pmf of X

  for any continuous variable X:
$\int p(x)\, dx = 1$, where p is the pdf of X
- $E[X + Y] = E[X] + E[Y]$
$E[aX + b] = aE[X] + b$
$Var[aX] = a^2\, Var[X]$
$Var[X] = E[X^2] - (E[X])^2$
$Cov(X, Y) = E[XY] - E[X]E[Y]$
- X, Y independent variables $\Rightarrow$
$Var[X + Y] = Var[X] + Var[Y]$
- X, Y independent variables $\Rightarrow$
$Cov(X, Y) = 0$, i.e. $E[XY] = E[X]E[Y]$

Advanced issues:

- For any vector of random variables, the covariance matrix is symmetric and positive semi-definite.

Ciortuz et al.'s exercise book: ch. 1, ex. 25

Exercises illustrating the above concepts/definitions and theoretical results/formulas, concentrating especially on:

- identifying in a given problem's text the underlying probabilistic distribution: either a basic one (e.g., Bernoulli, binomial, categorial, multinomial etc.), or one derived [by function composition or] by summation of identically distributed random variables
- computing probabilities
- computing means / expected values of random variables
- verifying the [conditional] independence of two or more random variables

Ciortuz et al.'s exercise book: ch. 1, ex. 9-16 [17-22], 46-55 [57-63], 64

Implementation exercises for advanced issues:

1. CMU, 2009 fall, Geoff Gordon, HW3, pr. 3
Implement *Linear Regression* and apply it to the task of predicting the level of PSA (Prostate Specific Agent) in prostate tissue, using a set of 8 variables (medical test results).[1]

2. CMU, 2014 fall, William Cohen, Ziv Bar-Jospeh, HW3, pr. 1
Implement *Logistic Regression* and apply it to the task of hand-written character recognition.[2]

[1] A somehow simpler exercise, CMU, 2009 spring, Ziv Bar-Joseph, HW1, pr. 4, uses linear regression on the compute the quantity of insulin to be injected into a patient based on his/her blood sugar level.

[2] A similar exercise, CMU, 2009 spring, Ziv Bar-Joseph, HW2, pr. 4.1-2, applies logistic regression (LR) on the *Breast Cancer* dataset [while pr. 4.3-4 compares LR with the Rosemblatt *perceptron* on this dataset].

◦ CMU, 2014 fall, W. Cohen, Z. Bar-Joseph, HW3, pr. 1.4-6
Implement *Multinomial Logistic Regression* and apply it to the *ORL Faces* dataset [while pr. 1.1-3,7 compares (M)LR with K-NN, Gaussian Naive Bayes and Gaussian Joint Bayes on this dataset].

3. CMU, 2010 fall, Ziv Bar-Jospeh, HW2, pr. 4
Study the regularization effect for logistic regression, using L2 and (especially) L1 norm, especially for *feature selection*. Work on the *communities and crime* dataset.