Online Advertising
Exercises

Exercise 1
Give an example of utility functions, a matching and item prices such that this outcome is envy-free (i.e. nobody wants to change items at the current prices) but not bidder optimal (i.e. there exist lower prices that also allow an envy-free outcome).

Exercise 2
Give an example of utility functions, a matching and item prices such that this outcome is both envy-free and seller optimal, i.e. any other envy-free outcome will have lower prices. Why are seller optimal solutions not as interesting in practice?

Exercise 3
For 10 million runs on random instances you observe that algorithm A has an average running time of 10ms with a standard deviation of 2ms, whereas algorithm B has an average running time of 11ms with a standard deviation of 3ms. Will a standard t-test for equality of means output a significance difference? Would the answer be different if there were only 10 runs?

Exercise 4
Suppose through a toolbar or browser plugin you know all of a user's browsing history. How could you use this knowledge to improve search results (i) for this user, and (ii) for users in general?

Exercise 5
Suppose you throw a (biased) coin 10 times and you observe "heads" once and "tails" nine times. What is the maximum likelihood estimate of the probability for the two events? Prove this. Now suppose you want to estimate the (unknown) mean for a Gaussian distribution with known variance. Prove that the MLE for the mean is the average.

Exercise 6
A scientist has a machine learning algorithm with a tuning parameter t. He splits his data set into 80% training data and 20% test data. For different values of t he trains his algorithm on the 80% of data and then evaluates the performance on the (unseen) 20% of data. In the end he chooses the t that gives the best performance on the test set. Is it guaranteed that for new data from the same distribution this value of t will perform best?

Exercise 7
Suppose you have access to all Twitter tweets. By studying the "retweet" (RT) of messages you want to find out key influencers in the network. How would you do this, i.e. what type of formula/equation/model would you come up with? How would you compare different notions of importance or influence? Which features of the message being retweeted or the individual users could help to improve the model? What are problems when it comes to evaluating model parameters?