582631 Introduction to Machine Learning, Fall 2016
Exercise set I
Model solutions

1.

(a) Let's solve the value of ϵ for which the upper bound for the considered probability equals α:

$$\begin{aligned}\alpha &= 2e^{-2n\epsilon^2}\\ e^{-2n\epsilon^2} &= \frac{\alpha}{2}\\ -2n\epsilon^2 &= \ln\alpha - \ln 2\\ \epsilon &= \pm\sqrt{\frac{\ln 2 - \ln\alpha}{2n}}.\end{aligned}$$

The length of the considered interval is $2n\epsilon$, and by plugging $\alpha = 0.05$ and $n = 10, 100, 1000$ into the formula we get:

n	10	100	1000
$2n\epsilon$	8.6	27.2	85.9

(b) Using the union bound and the previous exercise we get:

$$P(\bigcup_{i=1}^{k} A_i) \leq \sum_{i=1}^{k} P(A_i) \leq 2ke^{-2n\epsilon^2}.$$

Let's again solve the value of ϵ for which this upper bound for equals α:

$$\begin{aligned}\alpha &= 2ke^{-2n\epsilon^2}\\ \epsilon &= \pm\sqrt{\frac{\ln(2k) - \ln\alpha}{2n}}.\end{aligned}$$

By plugging $\alpha = 0.05$, and different values of n and k into the formula, we get the following interval lengths:

	n	10	100	1000
$k = 1$	$2n\epsilon$	8.6	27.2	85.9
$k = 10$	$2n\epsilon$	10.9	34.6	109.5
$k = 100$	$2n\epsilon$	12.9	40.7	128.8

Notice that while the width of the above interval, within which the number of correct predictions $\sum_i X_i$ is likely to be, grows with n at rate $\sqrt{n}$, the corresponding interval for the *proportion* of correct predictions $n^{-1}\sum_i X_i$ shrinks at rate $1/\sqrt{n}$:

	n	10	100	1000
$k = 1$	2ϵ	0.86	0.27	0.09
$k = 10$	2ϵ	1.09	0.35	0.11
$k = 100$	2ϵ	1.29	0.41	0.13

In summary, the observed accuracy (number of correct prediction divided by n) tends to get closer and closer to the true accuracy p as the sample size n grows. On the other hand, as the number of classifiers, k, is increased, the interval grows but as can be deduced from the formula for ϵ, the dependency on k is of the order $\sqrt{(\ln k)}$, which is very slow (as can also be seen in the above tables).