Intro to Statistical Learning Theory
Exercise 4

1) Prove the second inequality in lemma 3.5, lecture 7.
Hint: Look at the two case $a \geq b$ and $a < b$. consider the function $f(x) = KL(a||b) - \frac{(a-b)^2}{2b}$.

2) Suppose a PAC-Bayes algorithm returns posterior $Q(S)$ for any sample. Prove that prior that minimizes $P^* = \arg\min \mathbb{E}_{S\sim\mathcal{D}^m}[KL(Q||P)]$ is $\mathbb{E}_S[Q(S)]$.

3) Prove a compression bound theorem for SVM, based on the *unknown* number of support vectors. You can assume that the number of support vectors k is smaller then $m/2$.
Hint: Follow the same line as theorem 2.3 in lecture 5.

4) a) Prove that after T rounds of adaBoost, the fraction of training samples with margin at most θ is bounded by $\prod^t \sqrt{(1+2\gamma_t)^{(1+\theta)}(1-2\gamma_t)^{(1-\theta)}}$
hint: Prove first that $\exp(-y\sum \alpha_t h_t(x) + \theta\sum\alpha_t) \geq 1$ iff $yf(x) \leq \theta$

   b) Assume $\forall t, \gamma_t > \gamma$. Find for which values of θ, this bound decays exponentially.

5) Prove that given a weak learner we can produce a strong learner through boosting.