Intro to Statistical Learning Theory
Exercise 5

1) Prove the second inequality in lemma 3.5, lecture 1.
   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 every sample S. 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 that given a weak learner we can produce a strong learner through boosting.

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) Variational inference: In Bayesian inference we need to compute $p(y) = \int p(y|z)p(z)dz$. This can be hard to do, and sampling from $p(z)$ can be hard as well (so simple monte-carlo methods aren't practical). A solution is to replace $p(z)$ with $q(z)$ with we can sample easily from. Prove the "evidence lower bound"

$$\log p(y) = \log \int p(y|z)p(z)dz \geq \mathbb{E}_q[\log p(y|z)] - KL[q(z)||p(z)]$$

   i.e. the KL divergence bounds the penalty we pay for doing the expectation according to q instead of p.