Home Work Assignment 1
Due date: Nov. 21st

1. Given a vector of random variables $\mathbf{x} \in \mathbb{R}^n$ with covariance matrix Σ, $\Sigma_{ij} = \mathbb{E}[\mathbf{x}_i \mathbf{x}_j] - \mathbb{E}[\mathbf{x}_i]\mathbb{E}[\mathbf{x}_i]$.
   (a) If $\mathbf{w} \in \mathbb{R}^n$ and $y = \mathbf{w}^T \mathbf{x} = \sum_i^n \mathbf{w}_i \mathbf{x}_i$, show that $\mathbf{w}^T \Sigma \mathbf{w} = \mathrm{var}(y)$.
   (b) Show that Σ is positive semi-definite.

2. We define the KL divergence as $KL(p||q) = \mathbb{E}_{x\sim p}\left[\log(\frac{p(x)}{q(x)})\right] = \mathbb{E}_{x\sim p}\left[-\log(\frac{q(x)}{p(x)})\right]$ (infinity if q(x)=0 and $p(x) > 0$).
   (a) Prove that $KL(p||q) \geq 0$ and $KL(p||q) = 0$ if and only if $p = q$ (hint: Jensen inequality)
   (b) Show that minimizing the KL divergence $\theta^* = \arg\min_\theta KL(p||q_\theta)$ (note that it isn't symmetric so order matters) is the same as negative log-likelihood $\theta^* = \arg\max_\theta \mathbb{E}_p[-\log(q_\theta(x))]$
   (c) Show that if $p(x_1, ..., x_n) = \prod p_i(x_i)$ and $q(x_1, ..., x_n) = \prod q_i(x_i)$ are independent then $KL(p||q) = \sum_i KL(p_i||q_i)$.
   (d) Compute the KL divergence between two Gaussians $p(x) \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $q(x) \sim \mathcal{N}(\mu_2, \sigma_2^2)$.