HOMEWORK PROBLEMS 09, ANLY 561, FALL 2018
DUE 11/11/18

Readings: §5.9, 6.1, 6.2, 6.3 of **Goodfellow, Bengio, and Courville**

Exercises:

1. For any fixed nonzero $\mathbf{v} \in \mathbb{R}^n$ and any fixed $b \in \mathbb{R}$, the set of points $\mathbf{x} \in \mathbb{R}^n$ satisfying
$$\mathbf{v}^T\mathbf{x} - b = 0$$
is called an **affine hyperplane** of dimension $n-1$ in $\mathbb{R}^n$. Let $\mathbf{y} \in \mathbb{R}^n$ and find the solution $\mathbf{x}^*$ (in terms of $\mathbf{v}$, b, and $\mathbf{y}$) to the program
$$\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{y} - \mathbf{x}\|_2^2 \text{ subject to } \mathbf{v}^T\mathbf{x} - b = 0.$$
This $\mathbf{x}^*$ is called the **projection** of $\mathbf{y}$ onto the affine hyperplane determined by $\mathbf{v}$ and b.

2. Suppose $X \in M_{n,d}$ is a data matrix with rows from a dataset $\{\mathbf{x}_i\}_{i=1}^n \subset \mathbb{R}^d$. This problem illustrates the role of optimization in Principal Component Analysis.
   (a) Show that $\sum_{i=1}^n \|\mathbf{x}_i - c_i\mathbf{u}\|_2^2 = \|X - \mathbf{c}\mathbf{u}^T\|_{\text{Fro}}^2$, where $\|A\|_{\text{Fro}}^2 = \sum_{i=1}^m \sum_{j=1}^n A_{i,j}^2$ is the Frobenius norm of A for any $A = [A_{i,j}] \in M_{m,n}$, and $\mathbf{c} \in \mathbb{R}^n$.
   (b) Show that, for any $\mathbf{u} \in \mathbb{R}^d$ with $\|u\|_2^2 = 1$, $c_i^* = \mathbf{x}_i^T\mathbf{u}$ is the unique solution to $\min_{c_i\in\mathbb{R}} \|\mathbf{x}_i - c_i\mathbf{u}\|_2^2$.
   (c) Using part (a) and (b), explain why the optimization program
   $$\min_{\mathbf{u}\in\mathbb{R}^d, \mathbf{c}\in\mathbb{R}^n} \sum_{i=1}^n \|\mathbf{x}_i - c_i\mathbf{u}\|_2^2 \text{ subject to } \|\mathbf{u}\|_2^2 = 1$$
   can be solved if we solve the problem
   $$\min_{\mathbf{u}\in\mathbb{R}^d} \|X - X\mathbf{u}\mathbf{u}^T\|_{\text{Fro}}^2 \text{ subject to } \|\mathbf{u}\|_2^2 = 1$$
   (d) Show that $\|A\|_{\text{Fro}}^2 = \text{trace}(A^TA)$, where $\text{trace}(Q) = \sum_{i=1}^k Q_{i,i}$ is the sum of diagonal entries of any square matrix $Q = [Q_{i,j}] \in M_{k,k}$. Also, show that if $A, B \in M_{m,n}$, then $\text{trace}(A^TB) = \text{trace}(BA^T)$.
   (e) Use part (d) to show that the minimization program
   $$\min_{\mathbf{u}\in\mathbb{R}^d} \|X - X\mathbf{u}\mathbf{u}^T\|_{\text{Fro}}^2 \text{ subject to } \|\mathbf{u}\|_2^2 = 1$$
   and the maximization program
   $$\max_{\mathbf{u}\in\mathbb{R}^d} \mathbf{u}^TX^TX\mathbf{u} \text{ subject to } \|\mathbf{u}\|_2^2 = 1$$
   are equivalent.
   (f) Suppose $\mathbf{u}^*$ is a normalized eigenvector of X^TX corresponding to the largest eigenvalue of X^TX. Use part (e) and explain why $\mathbf{u}^*$ and $\mathbf{c}^* = X\mathbf{u}^*$ together solve the program
   $$\min_{\mathbf{u}\in\mathbb{R}^d, \mathbf{c}\in\mathbb{R}^n} \sum_{i=1}^n \|\mathbf{x}_i - c_i\mathbf{u}\|_2^2 \text{ subject to } \|\mathbf{u}\|_2^2 = 1$$