Machine Learning

Homework 2

Not collected, not graded.

1 High-dimensional PCA

We have seen the original PCA eigenvector problem as $\frac{1}{N}\mathbf{X}^T\mathbf{X}\mathbf{u}_i = \lambda_i \mathbf{u}_i$.
Using $\mathbf{v}_i = \mathbf{X}\mathbf{u}_i$, an equivalent eigenvector problem can be obtained: $\frac{1}{N}\mathbf{X}\mathbf{X}^T\mathbf{v}_i = \lambda_i \mathbf{v}_i$.
For $\mathbf{v}_i$ an eigenvector of the second problem, we find that $\mathbf{X}^T\mathbf{v}_i$ is an eigenvector of the original problem, with eigenvalue λ_i.

1. Assuming that $\mathbf{v}_i$ has unit length, show that $\mathbf{u}_i = \frac{1}{\sqrt{N\lambda_i}}\mathbf{X}^T\mathbf{v}_i$ also has unit length.

2 Probabilistic PCA

1. Let $\mathbf{x}$ be a D-dimensional random variable having Gaussian distribution $\mathcal{N}(\mathbf{x} \mid \mu, \Sigma)$, and consider the M-dimensional random variable given by $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$, where $\mathbf{A}$ is an $M \times D$ matrix. For $M = D$ and A non-singular, show that $\mathbf{y}$ is also Gaussian, and find expressions for its mean and covariance. (Linear algebra extra: What happens for $M < D$ and $M > D$?)

2. Consider two continuous variables $x, y \in \mathbb{R}$, with joint distribution $p(x,y)$. Let

$$\mathbb{E}_x[x \mid y](y) := \int p(x \mid y) x \, dx$$

   be the conditional expectation (as a function of y), and similarly

$$\mathrm{var}_x[x \mid y](y) := \int p(x \mid y)(x - \mathbb{E}_x[x \mid y])^2 \, dx$$

   the conditional variance. Show that

   (a) $\mathbb{E}[x] = \mathbb{E}_y[\mathbb{E}_x[x \mid y]]$, and

   (b) $\mathrm{var}[x] = \mathbb{E}_y[\mathrm{var}_x[x \mid y]] + \mathrm{var}_y[\mathbb{E}_x[x \mid y]]$ (Law of total variance).

3. Consider the latent variable $\mathbf{z} \in \mathbb{R}^D$ with zero-mean unit variance Gaussian distribution $\mathcal{N}(\mathbf{z} \mid 0, I_D)$. Let the observed variable $\mathbf{x} \in \mathbb{R}^M$ be given by the linear model

$$\mathbf{x} = W\mathbf{z} + \mu + \eta,$$

   for some $W \in \mathbb{R}^{M \times D}$, where η itself is zero-mean random Gaussian noise, $\mathcal{N}(\eta \mid 0, \sigma^2 I_M)$.

   Using the above results 2.(a) and (b), show that the predictive distribution

$$p(\mathbf{x}) = \int p(\mathbf{x} \mid \mathbf{z}) p(\mathbf{z}) \, d\mathbf{z}$$

   is itself Gaussian, with $p(\mathbf{x}) = \mathcal{N}(\mathbf{x} \mid \mu, C)$, for $C = WW^T + \sigma^2 I_M$.

4. For the same model, by using the results from §2.3.3 on p. 93, derive the posterior distribution to be

$$p(\mathbf{z} \mid \mathbf{x}) = \mathcal{N}(\mathbf{z} \mid M^{-1}W^T(\mathbf{x} - \mu), \sigma^2 M^{-1})$$

   with $M = W^T W + \sigma^2 I_M$.