Covariance and The Central Limit Theorem

1 The Covariance Matrix

Consider a probability density p on the real numbers. The mean and variance for this density is defined as follows.

$$\mu \;=\; \mathrm{E}_{\mathrm{x}\sim\mathrm{p}}\left[\mathrm{x}\right] = \int \mathrm{xp(x)}\ \mathrm{dx} \tag{1}$$

$$\sigma^2 \;=\; \mathrm{E}_{\mathrm{x}\sim\mathrm{p}}\left[(\mathrm{x}-\mu)^2\right] = \int (\mathrm{x}-\mu)^2\mathrm{p(x)}\ \mathrm{dx} \tag{2}$$

We now generalize mean and variance to dimension larger than 1. Consider a probability density p on R^D for $D \geq 1$. We can define a mean (or average) vector as follows.

$$\mu \;=\; \mathrm{E}_{\mathrm{x}\sim\mathrm{p}}\left[\mathrm{x}\right] = \int \mathrm{x}\ \mathrm{p(x)}\ \mathrm{dx} \tag{3}$$

$$\mu_i \;=\; \mathrm{E}_{\mathrm{x}\sim\mathrm{p}}\left[\mathrm{x_i}\right] = \int \mathrm{x_i}\ \mathrm{p(x)}\ \mathrm{dx} \tag{4}$$

We now consider a large sample x^1, ..., x^N drawn IID from p and we consider the vector average of these vectors.

$$\hat{\mu} = \frac{1}{N}\sum_{t=1}^{N} x^t$$

Assuming that the distribution mean μ is well defined (that the integral defining μ converge), we expect that as $N \to \infty$ we have that $\hat{mu}$ should approach the distribution average μ.

We now consider the generalization of variance. For variance we are interested in how the distribution varies around its mean. The variance of different dimensions can be different and, perhaps more importantly, the dimensions need not be independent. We define the *covariance matrix* by the following equation with i, j ranging from 1 to D.

$$\Sigma \;=\; \mathrm{E}_{\mathrm{x}\sim\mathrm{p}}\left[(\mathrm{x}-\mu)(\mathrm{x}-\mu)^{\mathrm{T}}\right] \tag{5}$$

$$\Sigma_{i,j} \;=\; \mathrm{E}_{\mathrm{x}\sim\mathrm{p}}\left[(\mathrm{x_i}-\mu_\mathrm{i})(\mathrm{x_j}-\mu_\mathrm{j})\right] \tag{6}$$

The definition immediately implies that Σ is symmetric, i.e., $\Sigma_{i,j} = \Sigma_{j,i}$. We also have that Σ is positive semidefinite meaning that $x\Sigma x \geq 0$ for all vectors x. We first give a general interpretation of Σ by noting the following.

$$
\begin{aligned}
u\Sigma v \;&=\; \mathrm{E}_{\mathrm{x}\sim\mathrm{p}}\left[\mathrm{u}(\mathrm{x}-\mu)(\mathrm{x}-\mu)^{\mathrm{T}}\mathrm{v}\right] & (7)\\
&=\; \mathrm{E}_{\mathrm{x}\sim\mathrm{p}}\left[(\mathrm{u}\cdot(\mathrm{x}-\mu))((\mathrm{x}-\mu)\cdot\mathrm{v})\right] & (8)\\
& & (9)
\end{aligned}
$$

So for fixed (nonrandom) vectors u and v we have that $u\Sigma v$ is the expected product of the two random variables $(x-\mu)\cdot u$ and $(x-\mu)\cdot v$. In particular we have that $u\Sigma u$ is the expected value of $((x-\mu)\cdot u)^2$. This implies that $u\Sigma u \geq 0$. A matrix satisfying this property for all u is called positive semi-definite. The covariance matrix is always both symmetric and positive semi-definite.

2 Multivariate Central Limit Theorem

We now consider the standard estimator $\hat{\mu}$ of μ where $\hat{\mu}$ is derived froma a sample x^1, ..., x^N drawn indpendently according to the density p.

$$\hat{\mu} = \frac{1}{N}\sum_{t=1}^{N} x^t \tag{10}$$

Note that $\hat{m}u$ can have different values for different samples — $\hat{\mu}$ is a random variable. As the sample size increases we expect that $\hat{\mu}$ becomes near μ and hence the length of the vector $\hat{\mu} - \mu$ goes to zero. In fact the length of this vector decreases like $1/\sqrt{N}$. In fact, the probability density for the quanity $\sqrt{N}(\hat{\mu} - \mu)$ converges to a multivariate Gaussian with zero mean covariance equal to the covariance of the density p.

Theorem 1 *For any density p on R^D with finite mean and positive semi-definite covariance, and any (measurable) subset U of R^D we have the following where μ and Σ are the mean and covariance matrix of the density p.*

$$\lim_{N \to \infty} \; \mathrm{P}\left[\sqrt{\mathrm{N}}(\hat{\mu} - \mu) \in \mathrm{U}\right] \;\; = \;\; \mathrm{P}_{\mathrm{x} \sim \mathcal{N}(0, \Sigma)}\left[\mathrm{x} \in \mathrm{U}\right]$$

$$\mathcal{N}(\nu, \Sigma)(x) \;\; = \;\; \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \nu)^T \Sigma^{-1}(x - \nu)\right)$$

Note that we need that the the covariance Σ is positive semi-definite, i.e., $x^T \Sigma x > 0$, or else we have $|\Sigma| = 0$ in which case the normalization factor is infinite. If $|\Sigma| = 0$ we simply work in the linear subspace spanned by the non-zero eigenvectors of Σ.

3   Eigenvectors

An eigenvector of a matrix A is a vector x such that $Ax = \lambda x$ where λ is a scalar called the eigenvalue of x. In general the compoenents of an eigenvector can be complex numbers and the eigenvalues can be complex. However, any symmetric positive semi-definite real-valued matrix Σ has real-valued orthogonal eigenvectors with real eigenvalues.

We can see that two eigenvectors with different eigenvalues must be orthogonal by observing the following for any two eigenvectors x and y.

$$\begin{aligned}
y\Sigma x \;\; &= \;\; \lambda_x(y \cdot x) \\
&= \;\; x\Sigma y \\
&= \;\; \lambda_y(x \cdot y)
\end{aligned}$$

So if $\lambda_x \neq \lambda_y$ then we must have $x \cdot y = 0$.

If two eigenvectors have the same eigenvalue then any linear combination of those eigenvectors are also eigenvectors with that same eigenvalue. So for a given eigenvalue λ the set of eigenvectors with eigenvalue λ forms a linear

subspace and we can select orthogonal vectors spanning this subspace. So there always exists an orhtonormal set of eigenvectors of Σ.

It is often convenient to work in an orthonormal coordinate system where the coordinate axes are eigenvectors of Σ. In this coordinte system we have that Σ is a diagonal matrix with $\Sigma_{i,i} = \lambda_i$, the eigenvalue of coordinate i.

In the coorindate system where the axes are the eigenvectors of Σ the multivariate Gaussian distribution has the following form where σ_i^2 is the eigenvalue of eigenvector i.

$$\mathcal{N}(\nu, \Sigma)(x) \;=\; \frac{1}{(2\pi)^{D/2} \prod_i \sigma_i} \exp\left(-\frac{1}{2} \sum_i \frac{(x_i - \nu_i)^2}{\sigma_i^2}\right) \tag{11}$$

$$=\; \prod_i \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x_i - \nu_i)^2}{2\sigma_i^2}\right) \tag{12}$$

4 Problems

1. Let p be the density on R^2 defined as follows.

$$p(x,y) \;=\; \begin{cases} \frac{1}{2} & \text{if } 0 \le x \le 2 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Give the mean and covariance matrix of this density. Drawn some iso-density contours of the Gaussian with the same mean and covariance as p.

2. Consider the following density.

$$p(x,y) \;=\; \begin{cases} \frac{1}{2} & \text{if } 0 \le x + y \le 2 \text{ and } 0 \le y - x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

Give the mean of the distribution and the eigenvectors and eigenvalues of the covariance matrix. Hint: draw the density.