

The Gaussian Distribution

CSC 411/2515

Adopted from PRML 2.3

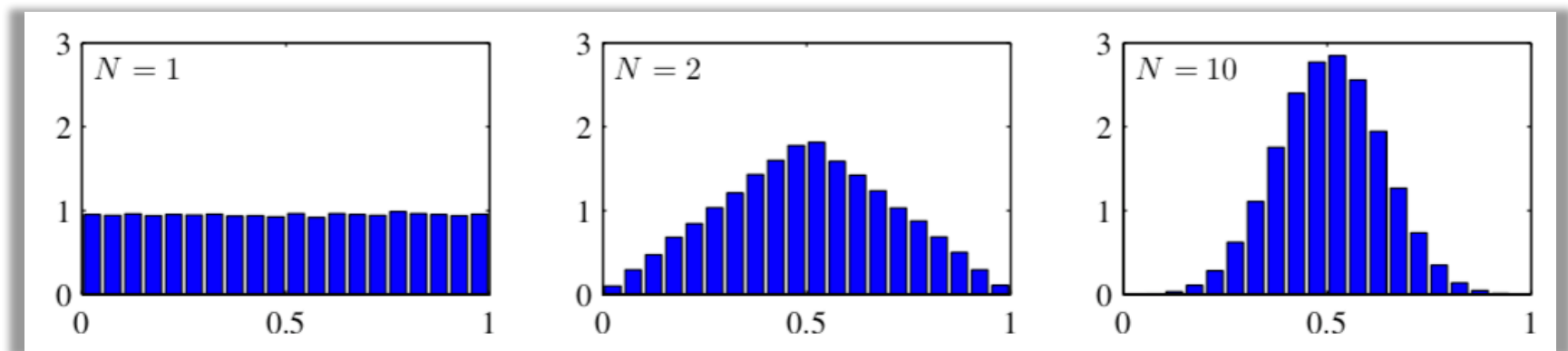
The Gaussian Distribution

- For a D -dimensional vector x , the multivariate Gaussian distribution takes the form:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - \mu)^\top \Sigma^{-1} (x - \mu) \right\}$$

- Motivations:

- Maximum of the entropy
- Central limit theorem



The Gaussian Distribution: Properties

- The law is a function of the **Mahalanobis** distance from x to μ :

$$\Delta^2 = (x - \mu)^\top \Sigma^{-1} (x - \mu)$$

- The expectation of x under the Gaussian distribution is:

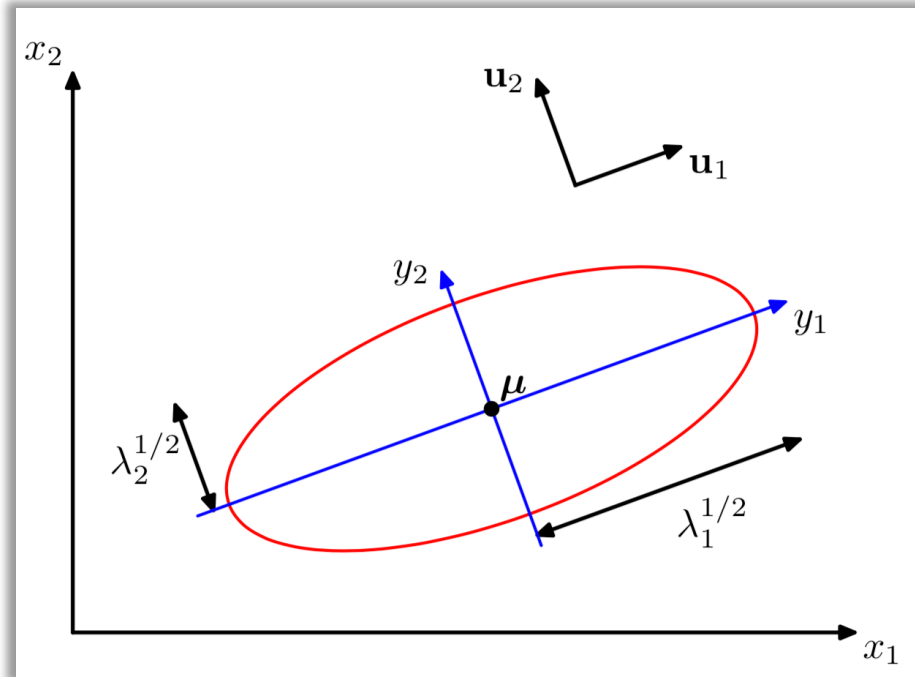
$$\mathbb{E}[x] = \mu$$

- The covariance matrix of x is:

$$\text{cov}[x] = \Sigma$$

The Gaussian Distribution: Properties

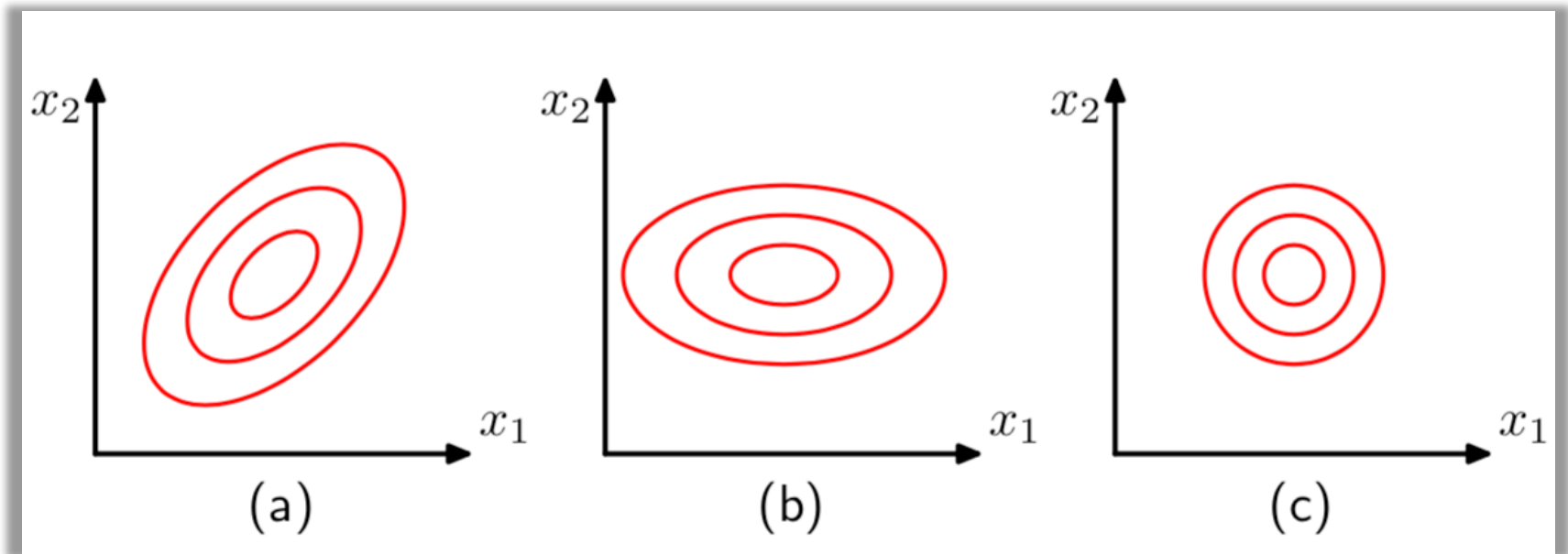
- The law (quadratic function) is constant on elliptical surfaces:



- λ_i are the eigenvalues of Σ
- u_i are the associated eigenvectors

The Gaussian Distribution: more examples

□ Contours of constant probability density:



- a) general form
- b) diagonal
- c) proportional to the identity matrix

Conditional Law

□ Given a Gaussian distribution $\mathcal{N}(x|\mu, \Sigma)$ with

$$x = (x_a, x_b)^\top, \quad \mu = (\mu_a, \mu_b)^\top$$

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}^{-1}$$

$$-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu) =$$

$$-\frac{1}{2}(x_a - \mu_a)^\top \Lambda_{aa}(x_a - \mu_a) - \frac{1}{2}(x_a - \mu_a)^\top \Lambda_{ab}(x_b - \mu_b)$$

$$-\frac{1}{2}(x_b - \mu_b)^\top \Lambda_{ba}(x_a - \mu_a) - \frac{1}{2}(x_b - \mu_b)^\top \Lambda_{bb}(x_b - \mu_b)$$

□ What's the conditional distribution $p(x_a|x_b)$?

Conditional Law

□ What's the conditional distribution $p(x_a|x_b)$?

$$-\frac{1}{2}(x - \mu)^\top \Sigma^{-1}(x - \mu) = -\frac{1}{2}x^\top \Sigma^{-1}x + x^\top \Sigma^{-1}\mu + \text{const}$$

- $\Sigma_{a|b}^{-1} = \Lambda_{aa}$
- $\Sigma_{a|b}^{-1}\mu_{a|b} = \Lambda_{aa}\mu_a - \Lambda_{ab}(x_b - \mu_b)$

□ Using the definition:

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}^{-1}$$

- $\Lambda_{aa} = (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}$
- $\Lambda_{ab} = -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1}$

Inverse partition identity:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{pmatrix} \quad M = (A - BD^{-1}C)^{-1}$$

Conditional Law

□ The conditional distribution $p(x_a|x_b)$ is a Gaussian with:

$$\mu_{a|b} = \mu_a + \Sigma_{ab}\Sigma_{bb}^{-1}(x_b - \mu_b)$$

$$\Sigma_{a|b} = \Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba}$$

□ The form using **precision matrix**:

$$\mu_{a|b} = \mu_a + \Lambda_{aa}\Lambda_{ab}^{-1}(x_b - \mu_b)$$

$$\Lambda_{a|b} = \Lambda_{aa}$$

Marginal Law

- The marginal distribution is given by:

$$p(x_a) = \int p(x_a, x_b) dx_b$$

- Picking out those terms that involve x_b , we have

$$-\frac{1}{2} x_b^\top \Lambda_{bb} x_b + x_b^\top m = -\frac{1}{2} (x_b - \Lambda_{bb}^{-1} m)^\top \Lambda_{bb} (x_b - \Lambda_{bb}^{-1} m) + \frac{1}{2} m^\top \Lambda_{bb}^{-1} m$$

$$m = \Lambda_{bb} \mu_b - \Lambda_{ba} (x_a - \mu_a)$$

- Integrate over x_b (unnormalized Gaussian)

$$\int \exp \left\{ -\frac{1}{2} (x_b - \Lambda_{bb}^{-1} m)^\top \Lambda_{bb} (x_b - \Lambda_{bb}^{-1} m) \right\} dx_b$$

- ✓ The integral is equal to the normalization term

Marginal Law

- After integrating over x_b , we pick out the remaining terms:

$$-\frac{1}{2}x_a^\top \Lambda_{aa}x_a + x_a^\top (\Lambda_{aa}\mu_a + \Lambda_{ab}\mu_b) + \frac{1}{2}m^\top \Lambda_{bb}^{-1}m + \text{const}$$

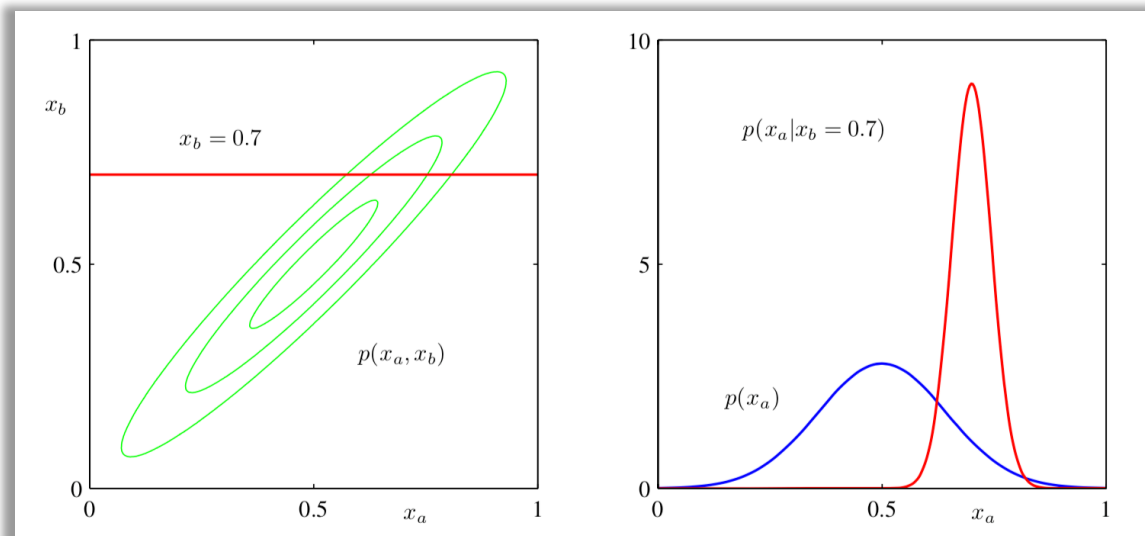
$$m = \Lambda_{bb}\mu_b - \Lambda_{ba}(x_a - \mu_a)$$

- The marginal distribution is a Gaussian with

$$\mathbb{E}[x_a] = \mu_a$$

$$\text{cov}[x_a] = \Sigma_{aa}$$

Short Summary



$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}^{-1}$$

□ Conditional distribution:

$$p(x_a | x_b) = \mathcal{N}(x_a | \mu_{a|b}, \Lambda_{aa}^{-1})$$
$$\mu_{a|b} = \mu_a - \Lambda_{aa}^{-1} \Lambda_{ab} (x_b - \mu_b)$$

□ Marginal distribution:

$$p(x_a) = \mathcal{N}(x_a | \mu_a, \Sigma_{aa})$$

Bayes' theorem for Gaussian variables

□ Setup:

$$p(x) = \mathcal{N}(x|\mu, \Lambda^{-1})$$

$$p(y|x) = \mathcal{N}(y|Ax + b, L^{-1})$$

□ What's the marginal distribution $p(y)$ and conditional distribution $p(x|y)$?

- ✓ How about first compute $p(z)$, where $z = (x, y)^\top$
- ✓ $p(z)$ is a Gaussian distribution, consider the log of the joint distribution

$$\begin{aligned}\ln p(z) &= \ln p(x) + \ln p(y|x) \\ &= -\frac{1}{2}(x - \mu)^\top \Lambda (x - \mu) \\ &\quad -\frac{1}{2}(y - Ax + b)^\top L (y - Ax + b) + \text{const}\end{aligned}$$

Bayes' theorem for Gaussian variables

- The same trick (consider the second order terms), we get

$$\mathbb{E}[z] = \begin{pmatrix} \mu \\ A\mu + b \end{pmatrix}$$

$$\text{cov}[z] = \begin{pmatrix} \Lambda^{-1} & \Lambda^{-1}A \\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A \end{pmatrix}$$

- We can then get $p(y)$ and $p(x|y)$ by **marginal and conditional laws!**

Maximum likelihood for the Gaussian

- Assume we have $X = (x_1, \dots, x_N)^\top$ in which the observations $\{x_n\}$ are assumed to be drawn independently from a multivariate Gaussian, the log likelihood function is given by

$$\ln p(X|\mu, \Sigma) = -\frac{ND}{2} \ln 2\pi - \frac{N}{2} \ln |\Sigma| - \frac{1}{2} \sum_{n=1}^N (x_n - \mu)^\top \Sigma^{-1} (x_n - \mu)$$

- Setting the derivative to zero, we obtain the solution for the maximum likelihood estimator:

$$\frac{\partial}{\partial \mu} \ln p(X|\mu, \Sigma) = \sum_{n=1}^N \Sigma^{-1} (x_n - \mu) = 0$$

$$\mu_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N x_n \quad \Sigma_{\text{ML}} = \frac{1}{N} \sum_{n=1}^N (x_n - \mu_{\text{ML}})(x_n - \mu_{\text{ML}})^\top$$

Maximum likelihood for the Gaussian

- The empirical mean is unbiased in the sense

$$\mathbb{E}[\mu_{\text{ML}}] = \mu$$

- However, the maximum likelihood estimate for the covariance has an expectation that is less than the true value:

$$\mathbb{E}[\Sigma_{\text{ML}}] = \frac{N-1}{N} \Sigma$$

- ✓ We can correct it by multiplying Σ_{ML} by the factor $\frac{N}{N-1}$

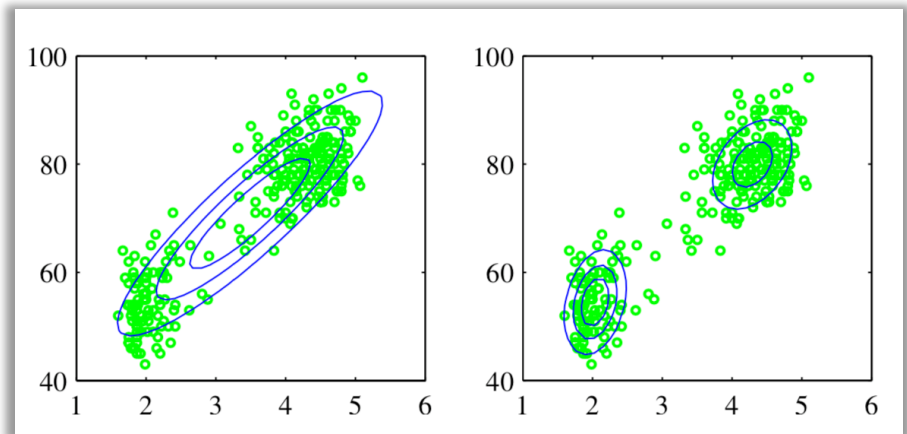
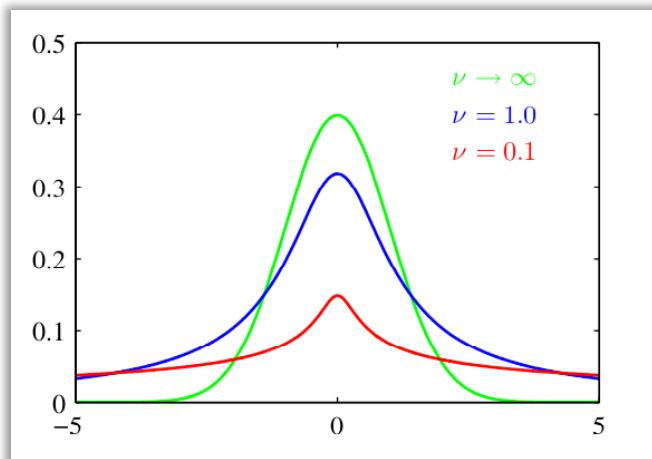
Conjugate prior for the Gaussian

- ❑ The maximum likelihood framework only gives point estimates for the parameters, we would like to have uncertainty estimation (confidence interval) for the estimation
 - ✓ Introducing prior distributions over the parameters of the Gaussian

- ❑ We would like the posterior $p(\theta|D) \propto p(\theta)p(D|\theta)$ has the same form as the prior (**Conjugate prior!**)
 - ✓ The conjugate prior for μ is a Gaussian
 - ✓ The conjugate prior for precision Λ is a Gamma distribution

The Gaussian Distribution: limitations

- ❑ A lot of parameters to estimate $D + \frac{D^2+D}{2}$: **structured approximation (e.g., diagonal variance matrix)**
- ❑ Maximum likelihood estimators are not robust to outliers: **Student's t-distribution (bottom left)**
- ❑ Not able to describe periodic data: **von Mises distribution**
- ❑ Unimodal distribution: **Mixture of Gaussian (bottom right)**



The Gaussian Distribution: frontiers

- Gaussian Process
- Bayesian Neural Networks
- Generative modeling (Variational Autoencoder)
-