# ISTA 421 + INFO 521
# Introduction to Machine Learning

**Lecture 8:**
**Continuous Probability, Gaussian Distribution**

**Clay Morrison**
claytonm@email.arizona.edu

18 September 2017

## Continuous Distributions:
Probability *Density* Functions (pdf)

## Continuous Spaces

- Outcome space is observation of a real value
  - E.g., person's height, temperature outside, etc…
- Example: a random variable *X* that can take on any value in [0,1] with equal probability
  - We say that *X* is **uniformly** distributed.

- Challenge: $P(X = x) = 0$
- Instead of assigning probabilities to points, we instead assign probabilities to regions (within some range or interval):

$P(x_1 \leq X \leq x_2)$ but not $P(X = x)$

3

## Probability Density Functions

To compute the probability that X lies in some range, we compute the definite integral of the density function:

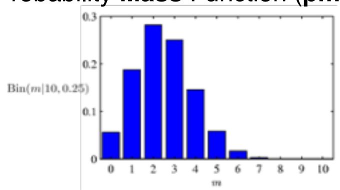
$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p(x)\, dx$$

# Probability **Density** Functions

To compute the probability that X lies in some range, we compute the definite integral of the density function:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p(x)\, dx$$

With discrete outcome spaces…
Probability **Mass** Function (**pmf**)

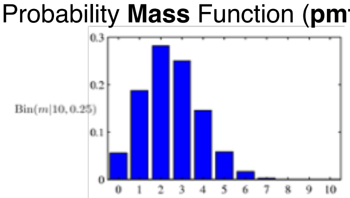



# Probability **Density** Functions

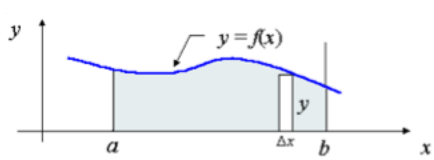
To compute the probability that X lies in some range, we compute the definite integral of the density function:

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p(x)\, dx$$

With discrete outcome spaces…
Probability **Mass** Function (**pmf**)




Probability **Density** Function (**pdf**)



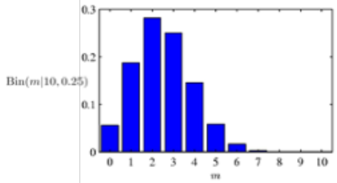
# Probability Density Functions

To compute the probability that X lies in some range, we compute the definite integral of the density function:

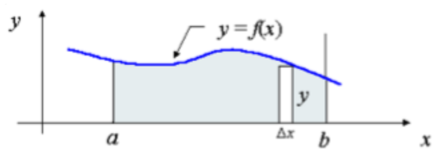
$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} p(x)\, dx$$

$$\underset{\text{density}}{\rho} = \frac{\overset{\text{mass}}{m}}{\underset{\text{volume}}{V}}$$

With discrete outcome spaces…
Probability **Mass** Function (**pmf**)




Probability **Density** Function (**pdf**)



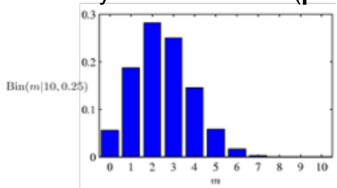

# Probability Density Functions

To compute the probability that X lies in some range, we compute the definite integral of the density function:

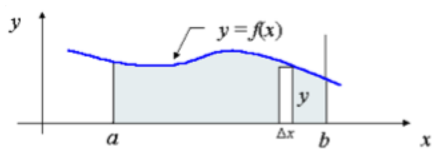
$$\underbrace{P(x_1 \leq X \leq x_2)}_{\text{mass}} = \int_{x_1}^{x_2} \underbrace{p(x)}_{\text{density}}\, \underbrace{dx}_{\text{“volume”}}$$

$$\underset{\text{density}}{\rho} = \frac{\overset{\text{mass}}{m}}{\underset{\text{volume}}{V}}$$

With discrete outcome spaces…
Probability **Mass** Function (**pmf**)




Probability **Density** Function (**pdf**)



## Probability Density Functions

To compute the probability that X lies in some range, we compute the definite integral of the density function:

$$\underbrace{P(x_1 \leq X \leq x_2)}_{\text{mass}} = \int_{x_1}^{x_2} \underbrace{p(x)}_{\text{density}}\, \underbrace{dx}_{\text{“volume”}}$$

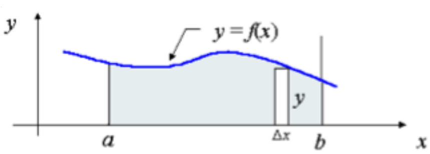
$$p : \mathbb{R} \mapsto \mathbb{R}$$

$$p(x) \geq 0$$

$$\int_{-\infty}^{\infty} p(x)\, dx = 1$$

Note: $p(x)$ can be > 1 !

Probability **Density** Function (**pdf**)




## Probability Density Functions

To compute the probability that X lies in some range, we compute the definite integral of the density function:

$$\underbrace{P(x_1 \leq X \leq x_2)}_{\text{mass}} = \int_{x_1}^{x_2} \underbrace{p(x)}_{\text{density}}\, \underbrace{dx}_{\text{“volume”}}$$

$p(\boldsymbol{x}) = p(x_0, x_1, ..., x_n)$ Probability vector is just a joint probability!

Joint $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_x) = \int_{x=x_1}^{x_2} \int_{y=y_1}^{y_2} p(x,y)\, dx\, dy$

Conditional $P(x_1 \leq X \leq x_2, Y = y) = \int_{x=x_1}^{x_2} p(x|Y=y)\, dx$

Marginalization $P(y) = \int_{x=x_1}^{x_2} p(x,y)\, dx$ (where $x_1 \leq X \leq x_2$ covers the sample space of *X*)

# Continuous Random Variables:

- Marginalization (summing out)

$$P(y) = \int_{x=x_1}^{x_2} p(x,y)\,dx$$

(where $x_1 \leq X \leq x_2$ describes the sample space of $X$)

- Expectations

$$\mathbf{E}_{p(x)}\{f(x)\} = \int f(x)p(x)\,dx$$

In many practical situations, not able to compute the integral (don't know the exact form of p(x), or it is impossible to integrate)

**Monte Carlo estimation** (if we can draw samples from $p(x)$)

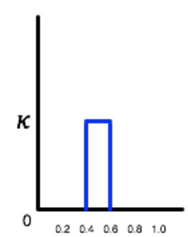
$$\mathbf{E}_{p(x)}\{f(x)\} \approx \frac{1}{S}\sum_{s=1}^{S} f(x)$$

The arithmetic mean!

# Uniform Distribution

**Example**: A random variable is **uniformly distributed** between 0.4 and 0.6, and never occurs outside that range

$$p(x) = \begin{cases} \kappa & x \in [0.4, 0.6] \\ 0 & \text{otherwise} \end{cases}$$




$$\int_{0.4}^{0.6} p(x)\,dx = \int_{0.4}^{0.6} \kappa\,dx = 1$$

$$= x\kappa\Big|_{0.4}^{0.6} = 0.6\kappa - 0.4\kappa$$

$$= 0.2\kappa = 1$$

$$\kappa = \frac{1}{0.2} = 5$$

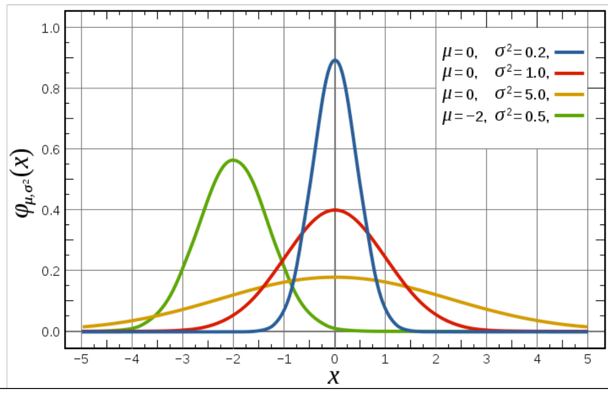
and thus

$$p(x) = \begin{cases} 5 & x \in [0.4, 0.6] \\ 0 & \text{otherwise} \end{cases}$$

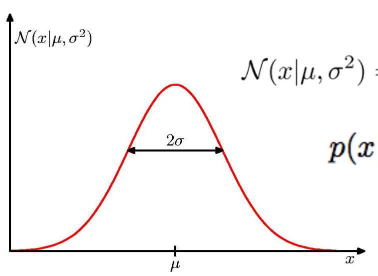
# Gaussian (Normal) Distribution

Univariate Gaussian (Normal) distribution

$$p(x|\mu,\sigma^2) = \mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

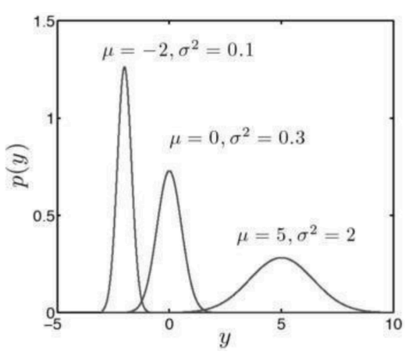



# Gaussian (Normal) Distribution

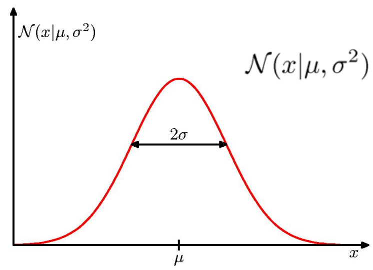



$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

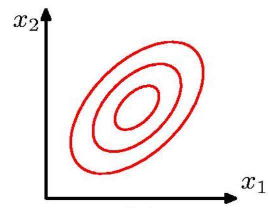
$$p(x|\mu,\sigma^2) = \mathcal{N}(x|\mu,\sigma^2) = \mathcal{N}(\mu,\sigma^2)$$



# Gaussian (Normal) Distribution

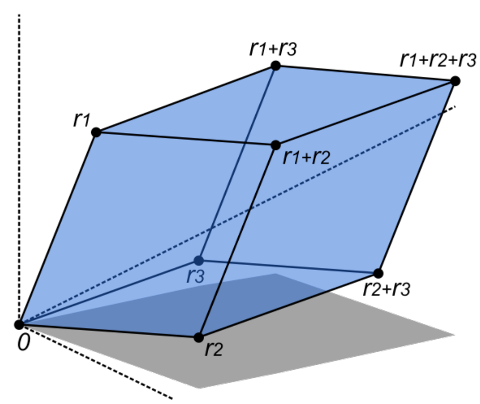



$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$




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

# A geometric interpretation of the determinant




The volume of this parallelepiped is the absolute value of the determinant of the matrix formed by the rows constructed from the vectors r1, r2 and r3

# Covariance Matrix in MV Guassian

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

$$cov(X,Y) = \mathbf{E}_{p(x,y)}\left\{\left(x-\mathbf{E}_{p(x)}\{x\}\right)\left(y-\mathbf{E}_{p(y)}\{y\}\right)\right\}$$

$$\Sigma = \begin{bmatrix} \mathrm{E}[(X_1-\mu_1)(X_1-\mu_1)] & \mathrm{E}[(X_1-\mu_1)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \mathrm{E}[(X_2-\mu_2)(X_1-\mu_1)] & \mathrm{E}[(X_2-\mu_2)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_2-\mu_2)(X_n-\mu_n)] \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{E}[(X_n-\mu_n)(X_1-\mu_1)] & \mathrm{E}[(X_n-\mu_n)(X_2-\mu_2)] & \cdots & \mathrm{E}[(X_n-\mu_n)(X_n-\mu_n)] \end{bmatrix}.$$

Covariance matrices are:
(1) Symmetric
(2) Positive semi-definite:
$z^T M z$ >= 0, for any non-zero vector z

# Sidebar: Positive Semi-definiteness

$z^T M z$ >= 0, for any non-zero vector z

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

# Sidebar: Positive Semi-definiteness

$z^T M z$ >= 0, for any non-zero vector z

$$z^{\mathrm{T}} I z = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2$$ Yes!

# Sidebar: Positive Semi-definiteness

$z^T M z$ >= 0, for any non-zero vector z

$$z^{\mathrm{T}} I z = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2$$ Yes!

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

# Sidebar: Positive Semi-definiteness

$z^T M z$ >= 0, for any non-zero vector z

$$z^{\mathrm{T}} I z = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2$$ Yes!

$$z^{\mathrm{T}} N z = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 \not> 0.$$


# Sidebar: Positive Semi-definiteness

$z^T M z$ >= 0, for any non-zero vector z

$$z^{\mathrm{T}} I z = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2$$ Yes!


$$z^{\mathrm{T}} N z = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 \not> 0.$$ No!

# Sidebar: Positive Semi-definiteness

$z^T M z$ >= 0, for any non-zero vector z

$$z^{\mathrm{T}} I z = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2$$ Yes!

$$z^{\mathrm{T}} N z = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 \not> 0.$$ No!

$$M = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$



# Sidebar: Positive Semi-definiteness

$z^T M z$ >= 0, for any non-zero vector z

$$z^{\mathrm{T}} I z = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2$$ Yes!

$$z^{\mathrm{T}} N z = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 \not> 0.$$ No!

$$M = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$


$$\begin{aligned} z^{\mathrm{T}} M z = (z^{\mathrm{T}} M) z &= \begin{bmatrix} (2a - b) & (-a + 2b - c) & (-b + 2c) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \\ &= 2a^2 - 2ab + 2b^2 - 2bc + 2c^2 \\ &= a^2 + (a - b)^2 + (b - c)^2 + c^2 \end{aligned}$$

# Sidebar: Positive Semi-definiteness

$z^T M z$ >= 0, for any non-zero vector z

$$z^{\mathrm{T}} I z = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = a^2 + b^2$$

Yes!

$$z^{\mathrm{T}} N z = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = -2 \not> 0.$$

No!

$$M = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

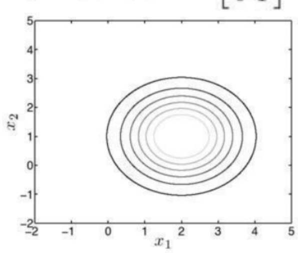
Yes!

$$z^{\mathrm{T}} M z = (z^{\mathrm{T}} M) z = \begin{bmatrix} (2a-b) & (-a+2b-c) & (-b+2c) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$
$$= 2a^2 - 2ab + 2b^2 - 2bc + 2c^2$$
$$= a^2 + (a-b)^2 + (b-c)^2 + c^2$$

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

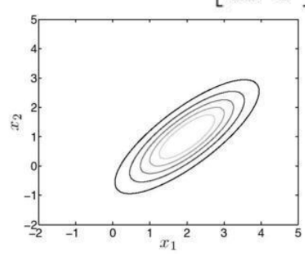
# Covariance in Gaussian Distribution

$\boldsymbol{\mu} = [2,\ 1]^{\mathrm{T}},\ \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$




(b)

$\boldsymbol{\mu} = [2,\ 1]^{\mathrm{T}},\ \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$




(d)

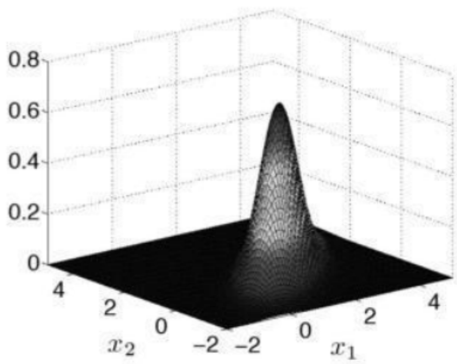
$$\sigma(x, y) = \mathbb{E}[(x - \mathbb{E}(x))(y - \mathbb{E}(y))]$$

$$\Sigma = \begin{bmatrix} \sigma(x,x) & \sigma(x,y) \\ \sigma(y,x) & \sigma(y,y) \end{bmatrix}$$

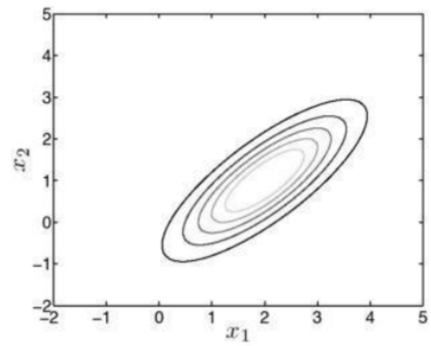
Also Note: **precision** (beta) used as inverse of variance: $\beta = \Sigma^{-1}$

# In 3D

$$\boldsymbol{\mu} = [2,\ 1]^{\mathsf{T}},\ \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

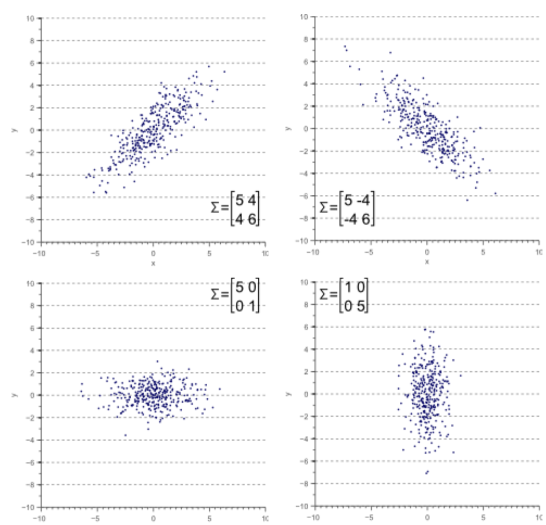



(c)




(d)

# More Examples (samples from $\boldsymbol{\Sigma}$)



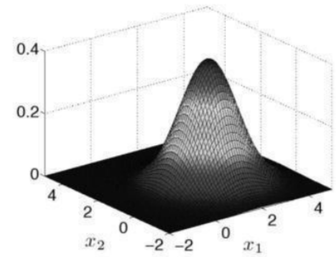
# The Multivariate Gaussian

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

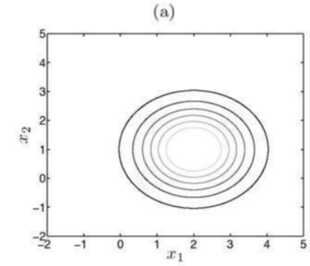
Example: $\boldsymbol{\mu} = [2,\ 1]^{\mathsf{T}},\ \boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$\begin{aligned}
p(\mathbf{x}) &= \frac{1}{(2\pi)^{D/2}|\mathbf{I}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{I}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\} \\
&= \frac{1}{(2\pi)^{D/2}|\mathbf{I}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathsf{T}}\mathbf{I}(\mathbf{x}-\boldsymbol{\mu})\right\} \\
&= \frac{1}{(2\pi)^{D/2}|\mathbf{I}|^{1/2}} \exp\left\{-\frac{1}{2}\sum_{d=1}^{D}(x_d-\mu_d)^2\right\} \\
&= \frac{1}{(2\pi)^{D/2}|\mathbf{I}|^{1/2}} \prod_{d=1}^{D} \exp\left\{-\frac{1}{2}(x_d-\mu_d)^2\right\} \\
&= \prod_{d=1}^{D} \frac{1}{(2\pi)^{1/2}} \exp\left\{-\frac{1}{2}(x_d-\mu_d)^2\right\}
\end{aligned}$$

Each term in the product is a univariate Gaussian!



(a)



(b)