State Estimation for Mobile Robotics

Michal Reinštein

Czech Technical University in Prague
Faculty of Electrical Engineering, Department of Cybernetics
Center for Machine Perception
`http://cmp.felk.cvut.cz/~reinsmic`,
`reinstein.michal@fel.cvut.cz`

Acknowledgement: V. Hlavac — Introduction to probability theory

Outline of the lecture:

- Probability rules
- Statistical moments
- Bayes theorem
- Maximum likelihood - MLE
- Maximum aposteriori - MAP
- Examples

Probability

is a function P, which assigns number from the interval $\langle 0, 1 \rangle$ to events and fulfils the following two conditions:

- $P(true) = 1$,
- $P\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} P(A_n)$, if the events $A_n,\ n \in \mathbb{N}$, are mutually exclusive.

From these conditions, it follows:

$P(false) = 0$, $\qquad P(\neg A) = 1 - P(A)$, $\qquad$ if $A \Rightarrow B$ then $P(A) \leq P(B)$.

Joint Probability

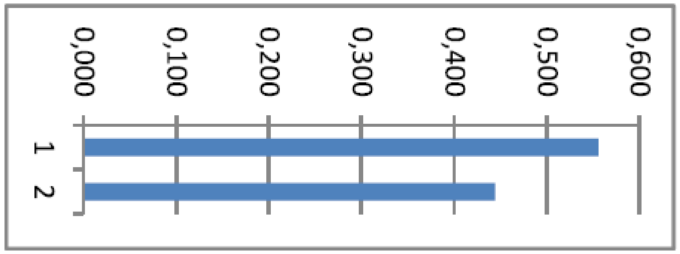
- The joint probability $P(A, B)$, also sometimes denoted $P(A \cap B)$, is the probability that events A, B co-occur.
- The joint probability is symmetric: $P(A, B) = P(B, A)$.

- Marginalization (the sum rule): $P(A) = \sum_B P(A, B)$ allows computing the probability of a single event A by summing the joint probabilities over all possible events B. The probability $P(A)$ is called the marginal probability.

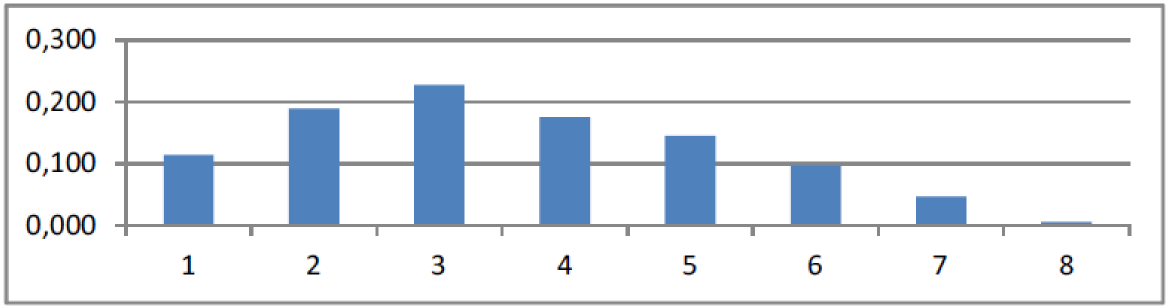
Marginalization

Orienteering competition example, participants									
Age	<= 15	16-25	26-35	36-45	46-55	56-65	66-75	>= 76	Sum
Men	22	36	45	33	29	21	12	2	200
Women	19	32	37	30	23	14	5	0	160
Sum	41	68	82	63	52	35	17	2	360

Orienteering competition example, frequency									
Age	<= 15	16-25	26-35	36-45	46-55	56-65	66-75	>= 76	Sum
Men	0,061	0,100	0,125	0,092	0,081	0,058	0,033	0,006	0,556
Women	0,053	0,089	0,103	0,083	0,064	0,039	0,014	0,000	0,444
Sum	0,114	0,189	0,228	0,175	0,144	0,097	0,047	0,006	1



Marginal probability P(sex)



Marginal probability P(Age_group)

Conditional Probability

- Let us have the probability representation of a system given by the joint probability $P(A, B)$.
- If an additional information is available that the event B occurred then our knowledge about the probability of the event A changes to

$$P(A|B) = \frac{P(A, B)}{P(B)},$$

which is the conditional probability of the event A under the condition B.

- The conditional probability is defined only for $P(B) \neq 0$.
- Product rule: $P(A, B) = P(A|B)\, P(B) = P(B|A)\, P(A)$.

Conditional Probability

- $P(true|B) = 1$, $P(false|B) = 0$.
- If $A = \bigcup_{n \in \mathbb{N}} A_n$ and events $A_1, A_2, \ldots$ are mutually exclusive then $P(A|B) = \sum_{n \in \mathbb{N}} P(A_n|B)$.
- Events A, B are independent $\Leftrightarrow P(A|B) = P(A)$.
- If $B \Rightarrow A$ then $P(A|B) = 1$.
- If $B \Rightarrow \neg A$ then $P(A|B) = 0$.

Conditional Probability

Example

Consider rolling a single dice. *What is the probability that the number higher than three comes up (event A) under the conditions that the odd number came up (event B)*?

$$\Omega = \{1,2,3,4,5,6\}\,, \quad A = \{4,5,6\}\,, \quad B = \{1,3,5\}$$

$$P(A) = P(B) = \frac{1}{2}$$

$$P(A,B) = P\left(\{5\}\right) = \frac{1}{6}$$

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

Independent Events

Events A, B are independent $\Leftrightarrow P(A,B) = P(A)\,P(B)$,
since independence means: $P(A|B) = P(A)$, $P(B|A) = P(B)$

Example

Rolling the dice once, events are: $A > 3$, event B is odd. Are A, B independent?

$$\Omega = \{1,2,3,4,5,6\}\,, \quad A = \{4,5,6\}\,, \quad B = \{1,3,5\}$$

$$P(A) = P(B) = \frac{1}{2}$$
$$P(A,B) = P\left(\{5\}\right) = \frac{1}{6}$$
$$P(A)\,P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

$P(A,B) \neq P(A)\,P(B) \Leftrightarrow$ The events are dependent.

Conditional Independence

Random events A, B are conditionally independent under the condition C, if

$$P(A, B|C) = P(A|C)\,P(B|C)\,.$$

Similarly, a conditional independence of more events, random variables, etc. is defined.

Definition of Bayes Theorem

$$\boxed{\mathrm{P(B|A)}} = \frac{\boxed{\mathrm{P(A|B)}}\; P(B)}{P(A)} \,,$$

where $\boxed{\mathrm{P(B|A)}}$ is the posterior probability and $\boxed{\mathrm{P(A|B)}}$ is the likelihood.

- This is a fundamental rule for machine learning (pattern recognition) as it allows to compute the probability of an output B given measurements A.
- The prior probability is $P(B)$ without any evidence from measurements.
- The likelihood $\boxed{\mathrm{P(A|B)}}$ evaluates the measurements given an output B. Seeking the output that maximizes the likelihood (*the most likely output*) is known as the maximum likelihood estimation (ML).
- The posterior probability $\boxed{\mathrm{P(B|A)}}$ is the probability of B after taking the measurement A into account. Its maximization leads to the maximum a-posteriori estimation (MAP).

Probability Rules

- The Product rule: $P(A, B) = P(A|B)\,P(B) = P(B|A)\,P(A)$
- The Sum rule: $P(B) = \sum_A P(A, B) = \sum_A P(B|A)\,P(A)$
- Random events A, B are independent $\Leftrightarrow P(A, B) = P(A)\,P(B)$,
- and the independence means: $P(A|B) = P(A)$, $P(B|A) = P(B)$
- A, B are conditionally independent $\Leftrightarrow P(A, B|C) = P(A|C)P(B|C)$
- The Bayes theorem:

$$P(A|B) = \frac{P(A,B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_A P(B|A)\,P(A)}$$

- General inference:

$$P(V|S) = \frac{P(V, S)}{P(S)} = \frac{\sum_{A,B,C} P(S, A, B, C, V)}{\sum_{V,A,B,C} P(S, A, B, C, V)}$$

Bayes Theorem

In Urban Search & Rescue (USAR), the ability of robots to reliably detect presence of a victim is crucial. How do we implement and evaluate this ability?

Example - Victim detection (1)

Assume we have a sensor S (e.g. a camera) and a computer vision algorithm that detects victims. We evaluated the sensor on ground truth data statistically:

- There is 20% chance of false negative detection (missed target).
- There is 10% chance of false positive detection.
- A priori probability of the victim presence V is 60%.

What is the probability that there is a victim if the sensor says no victim is detected?

Bayes Theorem

We express the sensor S measurements as a conditional probability of V:

$P(S\|V)$	$S = True$	$S = False$
$V = True$	0.8	0.2
$V = False$	0.1	0.9

Express the a priori knowledge as the probability:
$P(V = True) = 0.6$ and $P(V = False) = 1 - 0.6 = 0.4$

Express what-we-want: $P(V|S) =?$ given $S = False$ (not detecting a victim) and $V = True$ (but there is one).

Bayes Theorem

- Use the tools to express what-we-want in the terms of what-we-know:

$$P(V|S) = \frac{P(V,S)}{P(S)} = \frac{P(S|V)P(V)}{\sum_V P(S,V)} = \frac{P(S|V)P(V)}{\sum_V P(S|V)\,P(V)}$$

- Substitute $S = False$ and $V = True$ and sum over V to obtain:

$$P(V|S) = \frac{P(S = False|V = True)P(V = True)}{\sum_V P(S = False|V = True)\,P(V = True)} =$$

$$= \frac{0.2{\cdot}0.6}{0.2{\cdot}0.6{+}0.9{\cdot}0.4} = 0.25$$

Conclusion: if our sensors says there is no victim, we have **25%** chance of missing out someone! We need an additional sensor . . .

Bayes Theorem

In Urban Search & Rescue (USAR), the reliability is achieved through the sensor fusion: use the statistics to evaluate sensors and the probability theory to perform fusion.

Example - Victim detection (2)

Assume we have a sensor S as in the previous case and we add one more sensor T with the following properties:

- There is 5% chance of false negative detection (missed target).
- There is 5% chance of false positive detection.
- A priori probability of the victim presence is the same, V is 60%.

What is the probability that there is a victim if both sensors confirm its presence?

Bayes Theorem

We express the sensor T measurements as a conditional probability of V:

$P(T\|V)$	$T = True$	$T = False$
$V = True$	0.95	0.05
$V = False$	0.05	0.95

The a priori probability is the same:
$P(V = True) = 0.6$ and $P(V = False) = 1 - 0.6 = 0.4$

Express what-we-want: $P(V|S,T) =?$ given $S = True, T = True$ (both sensors see a victim) and $V = True$ (and there is one). Furthermore, we know that both sensors provide independent measurements with respect to each other.

Bayes Theorem

- Naive approach using joint probability: $P(S,T,V) = P(S,T|V)P(V)$
- Conditional independence: $P(S,T|V)P(V) = P(S|V)P(T|V)P(V)$
- Applying the tools:

$$P(V|S,T) = \frac{P(V,S,T)}{P(S,T)} = \frac{P(S|V)P(T|V)P(V)}{\sum_V P(V,S,T)} =$$

$$= \frac{P(S|V)P(T|V)P(V)}{\sum_V P(S|V)P(T|V)\,P(V)}$$

- Substitute: $S = True, T = True$, $V = True$ and sum over V to obtain:

$$= \frac{0.8{\cdot}0.95{\cdot}0.6}{0.8{\cdot}0.95{\cdot}0.6+0.1{\cdot}0.05{\cdot}0.4} = 0.9956$$

Conclusion: if both sensors confirm there is a victim, we have **99.56%** chance that there is a victim.

Random Variable

- The random variable is an arbitrary function $X : \Omega \to \mathbb{R}$, where Ω is a sample space.

- There are two basic types of random variables:

  - Discrete – a countable number of values.

    Examples: rolling a dice

    The discrete probability is given as: $P(X = a_i) = p(a_i)$, $i = 1, \ldots$, $\sum_i p(a_i) = 1$.

  - Continuous – values from some interval, i.e. infinite number of values.

    Example: the height persons.

    The continuous probability is given by the distribution function or the probability density function.

Distribution Function of a Random Variable

Distribution function of the random variable X is a function $F\colon X \to [0,1]$ defined as $F(x) = P(X \le x)$, where P is a probability.

Properties:

1. $F(x)$ is a non-decreasing function, i.e. $\forall$ pair $x_1 < x_2$ it holds $F(x_1) \le F(x_2)$.
2. $F(X)$ is continuous from the right, i.e. it holds $\lim\limits_{h\to 0^+} F(x+h) = F(x)$.
3. - It holds for every distribution function $\lim\limits_{x\to -\infty} F(x) = 0$ a $\lim\limits_{x\to\infty} F(x) = 1$. Written more concisely: $F(-\infty) = 0$, $F(\infty) = 1$.
   - If the possible values of $F(x)$ are from the interval (a,b) then $F(a) = 0$, $F(b) = 1$.

Continuous Distribution Function

- The distribution function F is called (absolutely) continuous if a nonnegative function f (probability density) exists and it holds

$$F(x) = \int_{-\infty}^{x} f(u)\,\mathrm{d}u \quad \text{for every } x \in X.$$

- The probability density fulfills

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1\,.$$

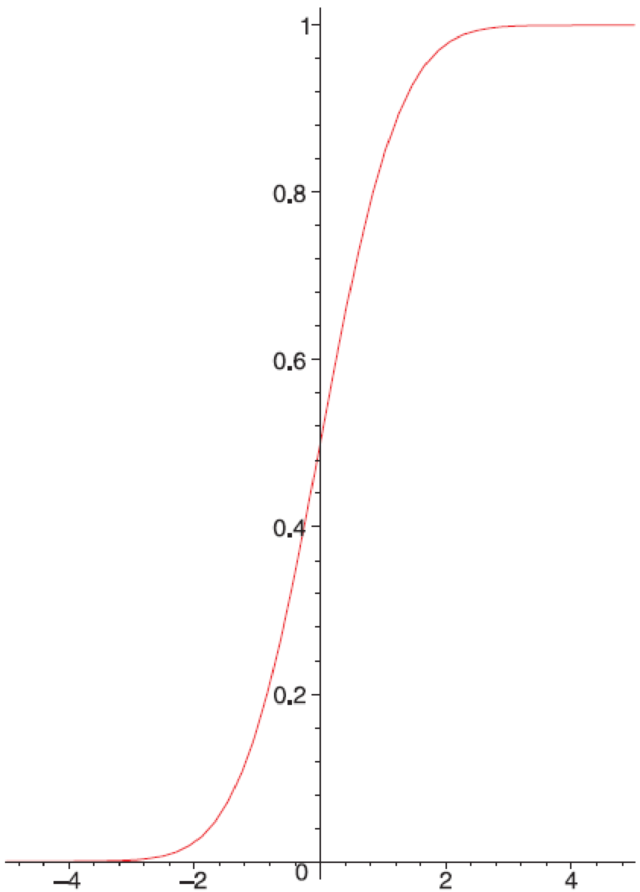
- If the derivative of $F(x)$ exists in the point x then $F'(x) = f(x)$.
- For $a, b \in \mathbb{R}$, $a < b$, it holds

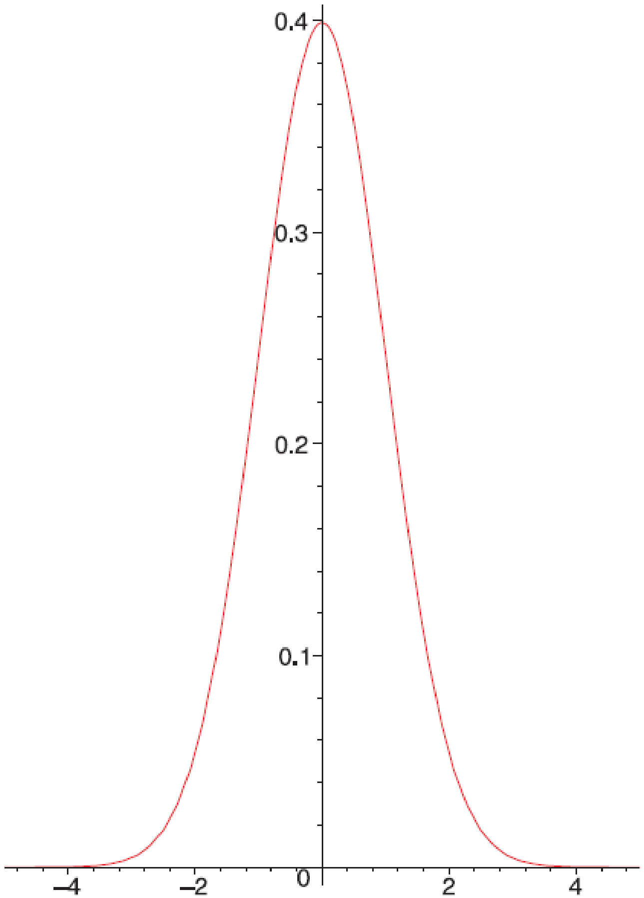
$$P(a < X < b) = \int_{a}^{b} f(x)\,\mathrm{d}x = F(b) - F(a)\,.$$

Normal Distribution

$$F(x)$$

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{-x^2}{2\sigma^2}}$$





Distribution function

Probability density

Expectation

- Expectation = the average of a variable under the probability distribution.
- Continuous definition: $E(x) = \int\limits_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x$.
- Discrete definition: $E(x) = \sum\limits_{x} x\, P(x)$.
- The expectation can be estimated from a N *number of samples* by $E(x) \approx \frac{1}{N} \sum\limits_{i} x_i$. The approximation becomes exact for $N \to \infty$.
- Expectation over multiple variables: $E_x(x, y) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} xy\, f(x, y)\, \mathrm{d}x\mathrm{d}y$
- Conditional expectation: $E(x|y) = \int\limits_{-\infty}^{\infty} x\, f(x|y)\, \mathrm{d}x$.

Statistical Moments

Continuous distribution	Discrete distribution

Expectation (mean)

$$E(x) = \int\limits_{-\infty}^{\infty} x\, f(x)\,\mathrm{d}x \qquad E(x) = \sum_x x\, P(x)$$

k-th (general) moment

$$E(x^k) = \int\limits_{-\infty}^{\infty} x^k\, f(x)\,\mathrm{d}x \qquad E(x) = \sum_x x^k\, P(x)$$

k-th central moment

$$E(x^k) = \int\limits_{-\infty}^{\infty} (x - E(x))^k\, f(x)\,\mathrm{d}x \qquad E(x) = \sum_x (x - E(x))^k\, P(x)$$

Dispersion (variance)

$$D(x) = \int\limits_{-\infty}^{\infty} (x - E(x))^2\, f(x)\,\mathrm{d}x \qquad E(x) = \sum_x (x - E(x))^2\, P(x)$$

Covariance

Mutual covariance σ_{xy} of two random variables X, Y is

$$\sigma_{xy} = E\left((X - \mu_x)(Y - \mu_y)\right)$$

Covariance matrix[1] Σ of n variables $X_1, \ldots, X_n$ is

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \ldots & \sigma_{1n}^2 \\ & \ddots & \\ \sigma_{n1}^2 & \ldots & \sigma_n^2 \end{bmatrix}$$

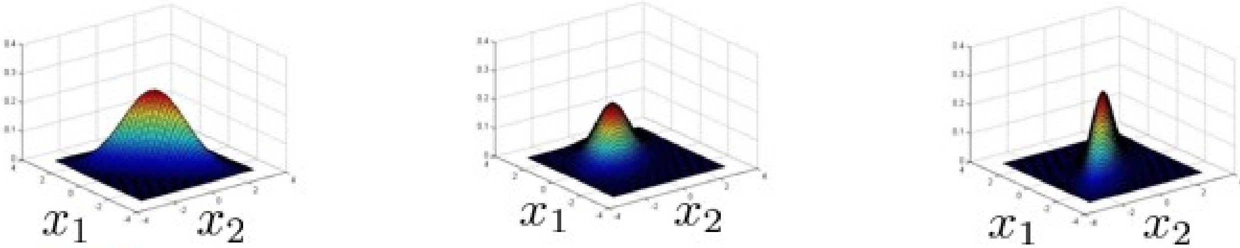
[1]Note: The covariance matrix is symmetric (i.e. $\Sigma = \Sigma^\top$) and positive-semidefinite (as the covariance matrix is real valued, the positive-semidefinite means that $x^\top M x \geq 0$ for all $x \in \mathbb{R}$).

Multivariate Normal distribution

Multivariate Gaussian (Normal) distribution

Parameters μ, Σ

$$p(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$



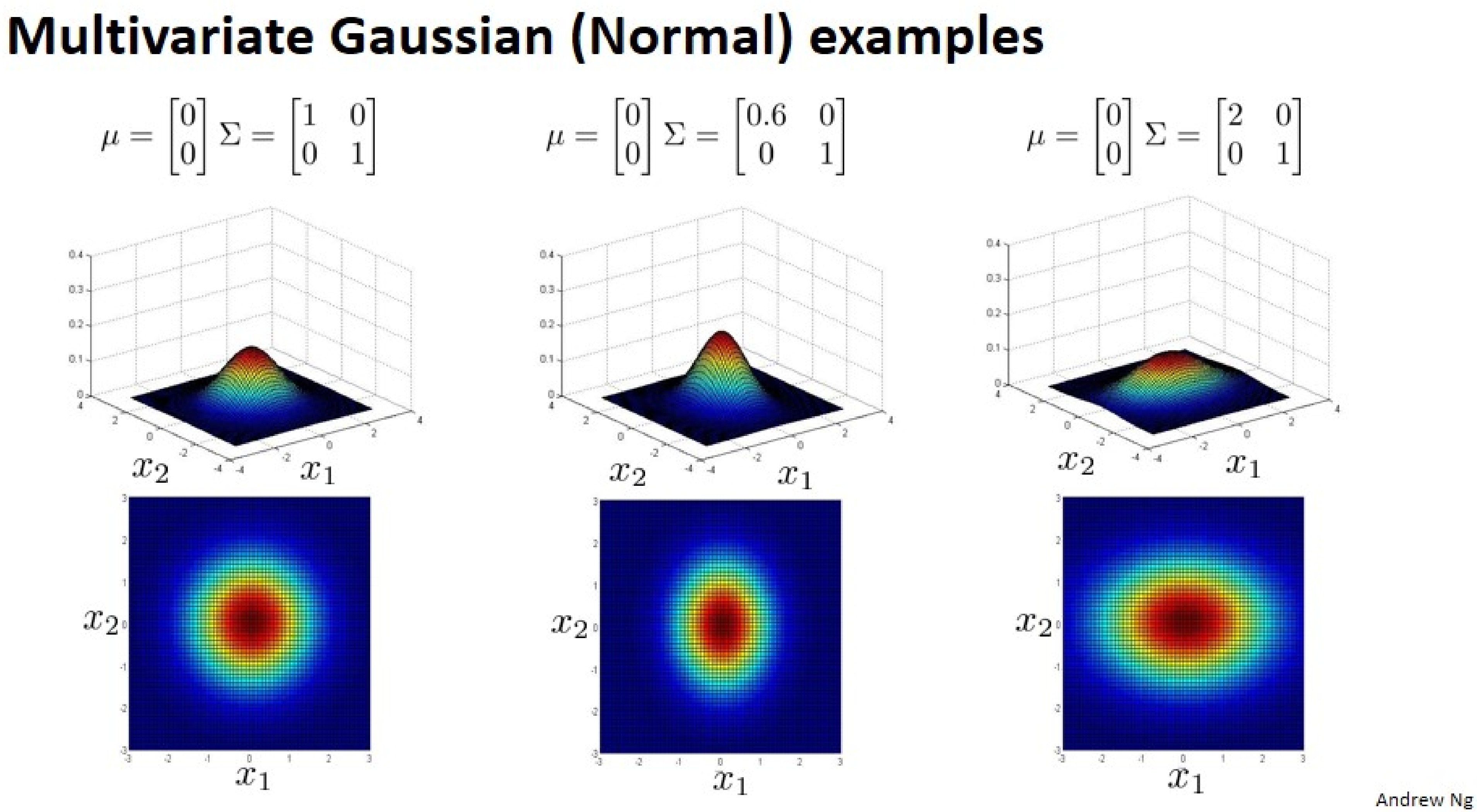
Parameter fitting:

Given training set $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x^{(i)} \qquad \Sigma = \frac{1}{m}\sum_{i=1}^{m}(x^{(i)}-\mu)(x^{(i)}-\mu)^T$$

Andrew Ng

Multivariate Normal distribution

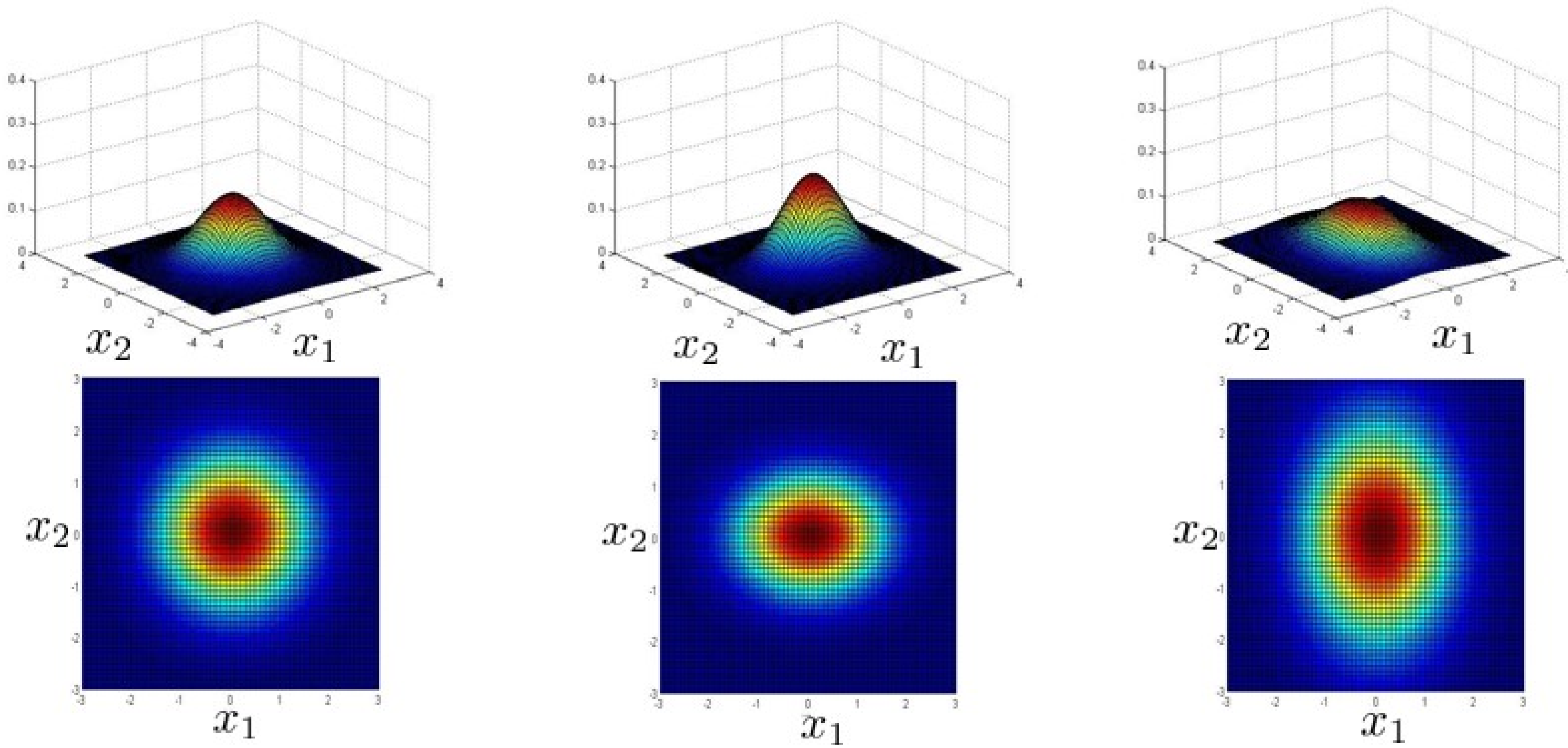


Source: Andrew Ng, Stanford University, Machine Learning course 2012, Lecture 15

Multivariate Normal distribution

Multivariate Gaussian (Normal) examples

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 0.6 \end{bmatrix} \qquad \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$



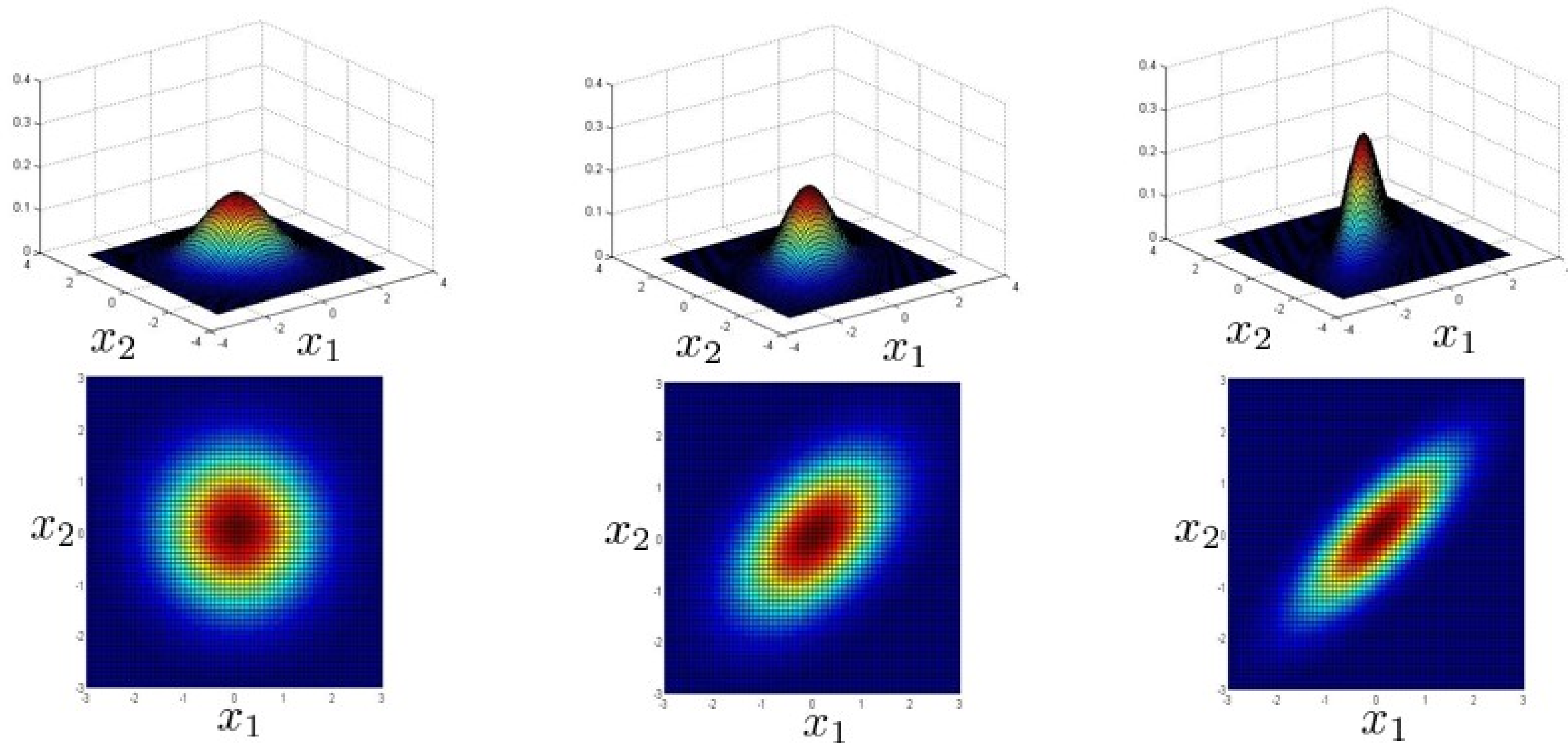
Andrew Ng

Source: Andrew Ng, Stanford University, Machine Learning course 2012, Lecture 15

Multivariate Normal distribution

Multivariate Gaussian (Normal) examples

$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$



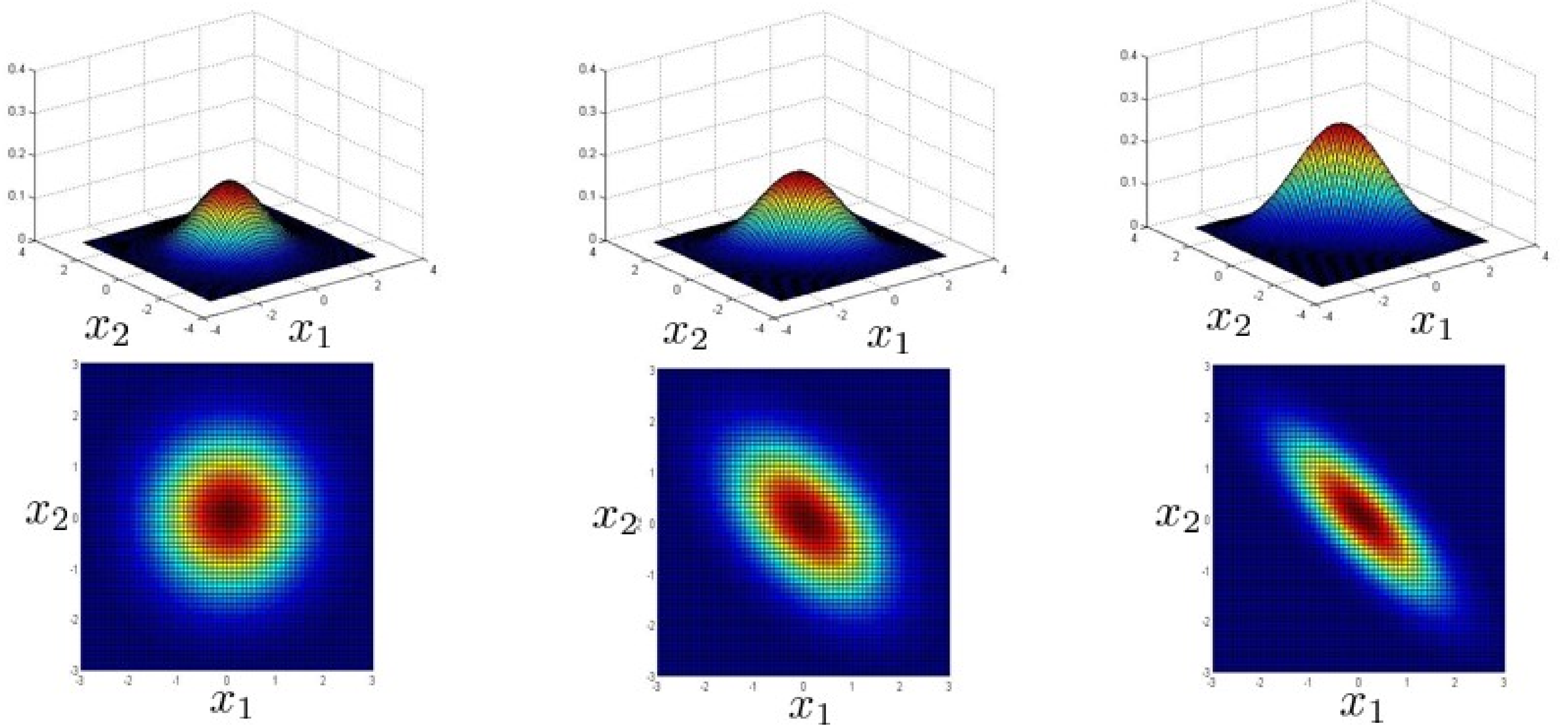
Andrew Ng

Source: Andrew Ng, Stanford University, Machine Learning course 2012, Lecture 15

Multivariate Normal distribution

Multivariate Gaussian (Normal) examples

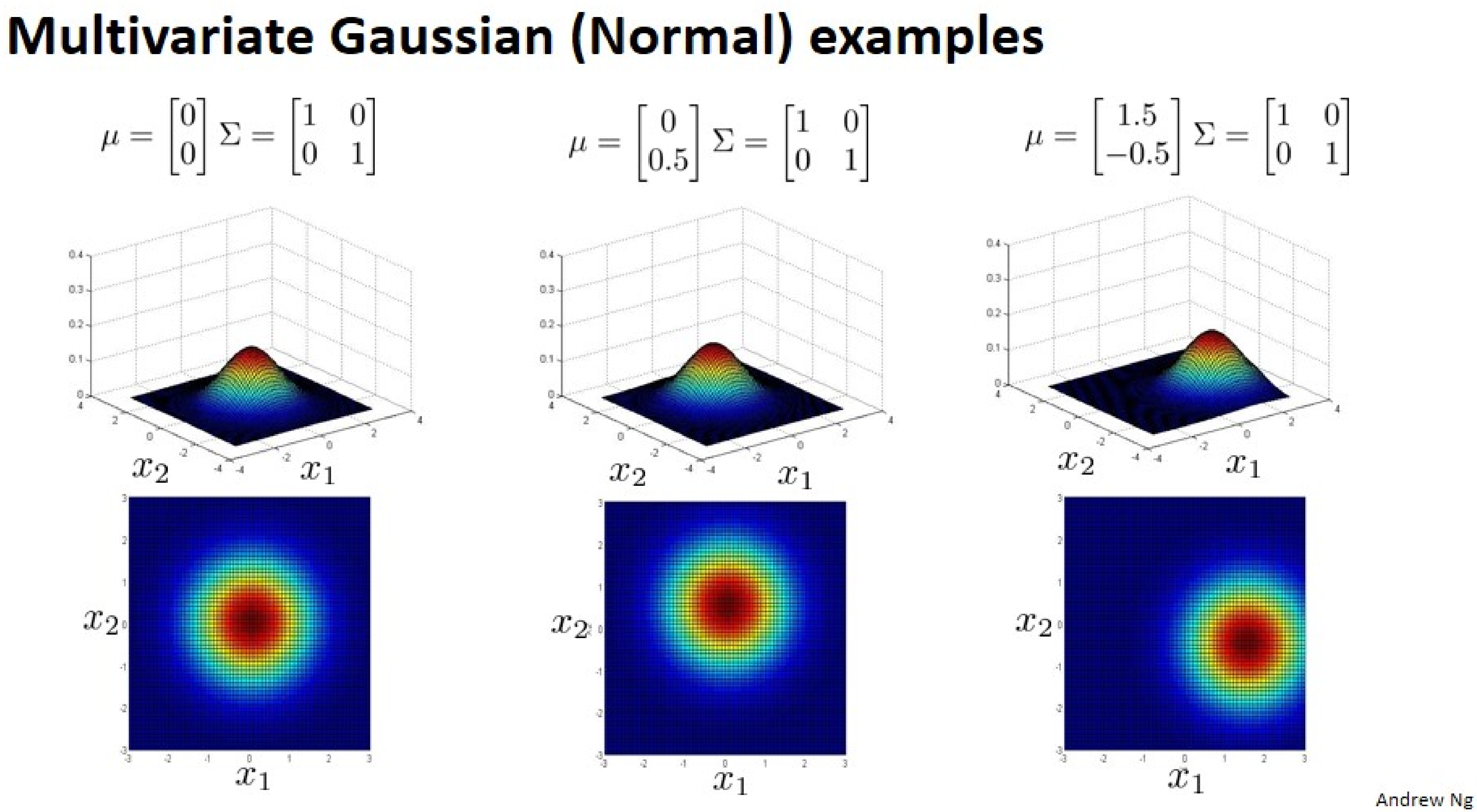
$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Sigma = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$



Andrew Ng

Source: Andrew Ng, Stanford University, Machine Learning course 2012, Lecture 15

Multivariate Normal distribution



Source: Andrew Ng, Stanford University, Machine Learning course 2012, Lecture 15

MLE - Maximum Likelihood Estimation

- The likelihood $\mathcal{L}(\mathbf{x})$ is the conditional probability $p(\mathbf{z}|\mathbf{x})$ of the measurements[2] $\mathbf{z}$ given a particular true value of $\mathbf{x}$.
- If the distribution is Gaussian and observations $\mathbf{z}$ are measured, the likelihood $\mathcal{L}(\mathbf{x})$ is a function only of $\mathbf{x}$.
- How do we obtain MLE? Knowing the distribution of $\mathcal{L}(\mathbf{x})$ and measurements $\mathbf{z}$, then $\mathbf{x}$ is varied until the maximum of the distribution is found:

$$\boxed{\hat{\mathbf{x}}_{MLE} = \operatorname*{argmax}_{x} p(\mathbf{z}|\mathbf{x})}$$

[2]Note: The likelihood is a function of $\mathbf{x}$ but it is not a probability distribution over $\mathbf{x}$, it would be incorrect to refer to it as the *likelihood of the data.*

MLE - Maximum Likelihood Estimation

Example - Sonar MLE (1)

Suppose we have two independent sonar measurements z_1, z_2 of a position x. The sensors are modeled both in the same way as $p(z_i|x) = \mathcal{N}(x, \sigma^2)$.

- Since the two sensors are independent the likelihood is:

$$\mathcal{L}(x) = p(z_1, z_2|x) = p(z_1|x)p(z_2|x)$$

- and since the sensors are Gaussian[3]:

$$\mathcal{L}(x) \sim e^{-\frac{(z_1-x)^2}{2\sigma^2}} \times e^{-\frac{(z_2-x)^2}{2\sigma^2}} = e^{-\frac{(z_1-x)^2+(z_2-x)^2}{2\sigma^2}}$$

[3] Note: we ignore the irrelevant normalization constant.

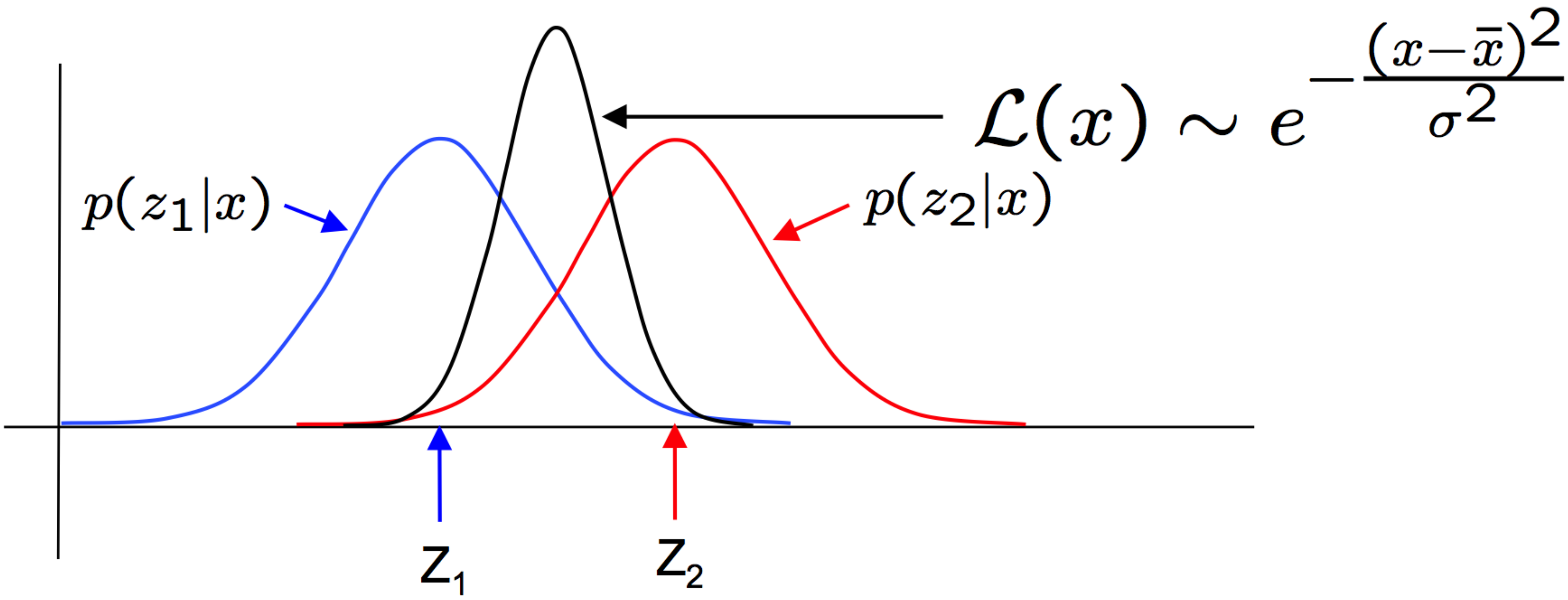
Example - Sonar MLE (2)

- We can express the negative log likelihood as follows:

$$-\ln \mathcal{L}(x) = \frac{(z_1 - x)^2 + (z_2 - x)^2}{2\sigma^2} = \frac{2x^2 - 2x(z_1 + z_2) + z_1^2 + z_2^2}{2\sigma^2}$$

- We redefine the MLE task to: $\hat{\mathbf{x}}_{\text{MLE}} = \underset{x}{\operatorname{argmin}} \; -\ln \mathcal{L}(x)$
- We minimize by differentiating w.r.t. x and setting equal to 0,
- which leads to: $\hat{\mathbf{x}}_{\text{MLE}} = \frac{z_1+z_2}{2} = \overline{x}$

Example - Sonar MLE (3)



MLE - Maximum Likelihood Estimation

Example - Sonar MLE (4)

Suppose we have two independent sonar measurements z_1, z_2 of a position x, but each sensor has a different model: $p(z_1|x) = \mathcal{N}(x, \sigma_1^2)$ and $p(z_2|x) = \mathcal{N}(x, \sigma_2^2)$.

- Again, the two sensors are independent and the likelihood is:

$$\mathcal{L}(x) = p(z_1, z_2|x) = p(z_1|x)p(z_2|x) \rightarrow \mathcal{L}(x) \sim e^{-\frac{(z_1-x)^2}{2\sigma_1^2}} \times e^{-\frac{(z_2-x)^2}{2\sigma_2^2}}$$

- We express the negative log likelihood:

$$-\ln \mathcal{L}(x) = 0.5(\sigma_1^{-2}(z_1 - x)^2 + \sigma_2^{-2}(z_2 - x)^2) + \text{const}$$

- and we minimize it by differentiating w.r.t. to x and setting to 0:

$$\hat{\mathbf{x}}_{\text{MLE}} = \frac{\sigma_1^{-2} z_1 + \sigma_2^{-2} z_2}{\sigma_1^{-2} + \sigma_2^{-2}}, \;\; \hat{\sigma}_{\text{MLE}}^{-2} = \sigma_1^{-2} + \sigma_2^{-2}$$

Example - Sonar MLE (5)

Now, assume we tested the sensors and we identified their variances of the measurements, such that: $p(z_1|x) \sim \mathcal{N}(x, 10^2)$ and $p(z_2|x) \sim \mathcal{N}(x, 20^2)$. What will be the MLE for these sensor readings $z_1 = 130$ and $z_2 = 170$?

$$\hat{\mathbf{x}}_{\text{MLE}} = \frac{130/10^2 + 170/20^2}{1/10^2 + 1/20^2} = 138$$

$$\hat{\sigma}_{\text{MLE}} = \frac{1}{\sqrt{1/10^2 + 1/20^2}} = 8.94$$

Conclusion: the ML estimate is closer to the more confident measurement.

MAP - Maximum A-Posteriori Estimation

- In many cases, we already have some prior (expected) knowledge about the random variable $\mathbf{x}$, i.e. the parameters of its probability distribution $p(\mathbf{x})$.
- With the Bayes rule, we go from prior to a-posterior knowledge about $\mathbf{x}$, when given the observations $\mathbf{z}$:

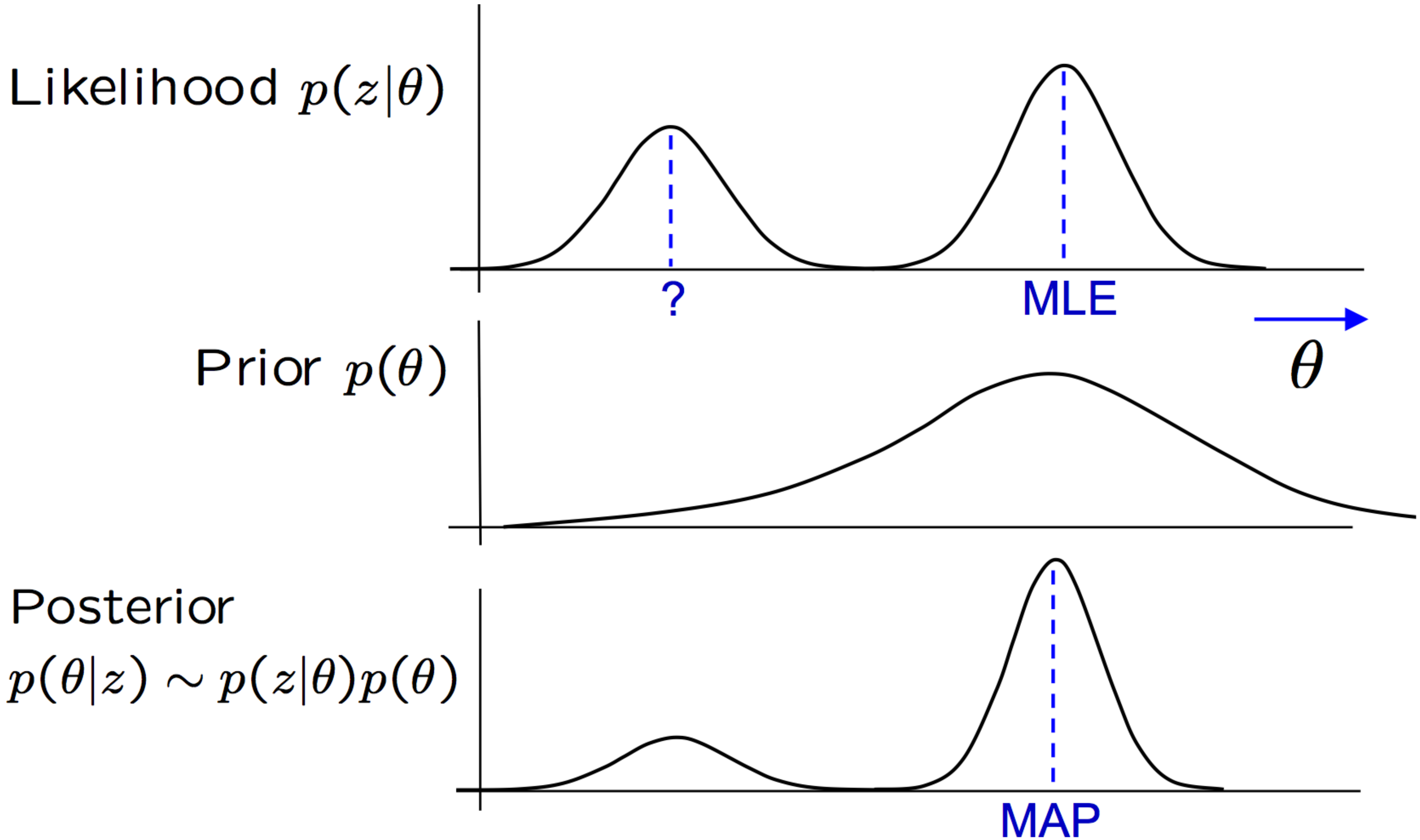
$$p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{z})} = \frac{\text{likelihood}\times\text{prior}}{\text{normalizing constant}} \sim C \times p(\mathbf{z}|\mathbf{x})p(\mathbf{x})$$

- Given an observation $\mathbf{z}$, a likelihood function $p(\mathbf{z}|\mathbf{x})$ and prior distribution $p(\mathbf{x})$ on $\mathbf{x}$, the maximum a posteriori estimator MAP finds the value of $\mathbf{x}$ which maximizes the posterior distribution $p(\mathbf{x}|\mathbf{z})$:

$$\hat{\mathbf{x}}_{\text{MAP}} = \operatorname*{argmax}_{x} p(\mathbf{z}|\mathbf{x})p(\mathbf{x})$$

MAP - Maximum A-Posteriori Estimation

Example - Application of MAP to a random variable of θ



MAP - Maximum A-Posteriori Estimation

Example - Sonar MAP (1)

Suppose we again have two independent sonar measurements z_1, z_2 of a position x, and each sensor modeled as: $p(z_1|x) = \mathcal{N}(x, \sigma_1^2)$ and $p(z_2|x) = \mathcal{N}(x, \sigma_2^2)$.

- The joint likelihood is defined as:

$$\mathcal{L}(x) = p(z_1, z_2|x) = p(z_1|x)p(z_2|x).$$

- In addition, we also have a prior (expected) information about x:

$$p(x) \sim \mathcal{N}(x_{prior}, \sigma_{prior}^2).$$

- The posterior probability density is given by a Gaussian distribution:

$$p(x|z_1, z_2) \sim p(z_1, z_2|x)p(x) \sim \mathcal{N}(x_{pos}, \sigma_{post}^2)$$

MAP - Maximum A-Posteriori Estimation

Example - Sonar MAP (2)

- Using the same approach as for deriving the MLE, the mean of the posteriori distribution of MAP is obtained as:

$$x_{post} = \frac{\sigma_1^{-2} z_1 + \sigma_2^{-2} z_2 + \sigma_{prior}^{-2} x_{prior}}{\sigma_1^{-2} + \sigma_2^{-2} + \sigma_{prior}^{-2}} = \hat{\mathbf{x}}_{\mathrm{MAP}}$$

- and the variance is:

$$\sigma_{post}^{-2} = \sigma_1^{-2} + \sigma_2^{-2} + \sigma_{prior}^{-2} = \hat{\sigma}_{\mathrm{MAP}}^{-2}$$

Example - Sonar MAP (3)

We assume the same sensors as in the previous example $p(z_1|x) \sim \mathcal{N}(x, 10^2)$ and $p(z_2|x) \sim \mathcal{N}(x, 20^2)$, but now consider a prior (expected) knowledge[4] $p(x) \sim \mathcal{N}(x_{prior} = 150, \sigma^2_{prior} = 30^2)$. What will be the MAP for these sensor readings $z_1 = 130$ and $z_2 = 170$?

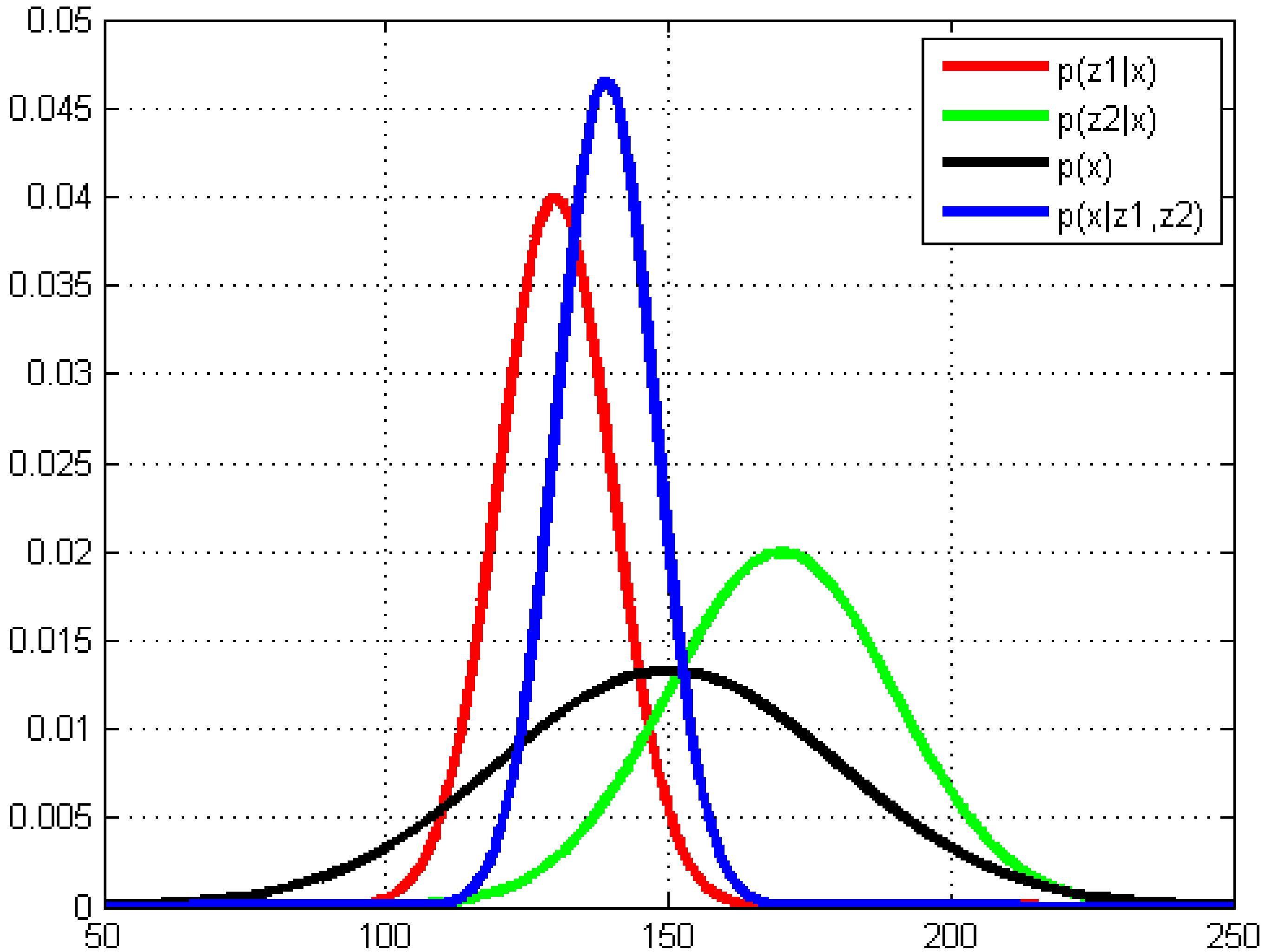
$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \frac{130/10^2 + 170/20^2 + 150/30^2}{1/10^2 + 1/20^2 + 1/30^2} = 139.04$$

$$\hat{\sigma}_{\mathrm{MAP}} = \frac{1}{\sqrt{1/10^2 + 1/20^2 + 1/30^2}} = 8.57$$

[4] Note: The prior knowledge is obtained for example statistically or from a datasheet.

MAP - Maximum A-Posteriori Estimation

Example - Sonar MAP (5)



What is the relationship between MLE and MAP?

The relationship between MLE and MAP is the update rule:

$$\hat{\mathbf{x}}_{\mathrm{MAP}} = \frac{\sigma_{prior}^{-2} x_{prior} + \sigma_{lik}^{-2} \hat{\mathbf{x}}_{\mathrm{MLE}}}{\sigma_{prior}^{-2} + \sigma_{lik}^{-2}} = \boxed{x_{prior} + \frac{\sigma_{prior}^{2}}{\sigma_{prior}^{2} + \sigma_{lik}^{2}} \left(\hat{\mathbf{x}}_{\mathrm{MLE}} - x_{prior}\right)}$$

- We can see that the prior acts as an additional sensor.
- If $\hat{\mathbf{x}}_{\mathrm{MLE}} = x_{prior}$ then $\hat{\mathbf{x}}_{\mathrm{MAP}}$ is unchanged by prior but variance decreases.
- If $\sigma_{lik} >> \sigma_{prior}$ then $\hat{\mathbf{x}}_{\mathrm{MAP}} \approx x_{prior}$ (noisy sensor!).
- If $\sigma_{prior} >> \sigma_{lik}$ then $\hat{\mathbf{x}}_{\mathrm{MAP}} \approx \hat{\mathbf{x}}_{\mathrm{MLE}}$ (weak prior knowledge!).