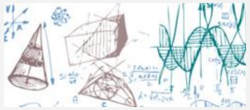
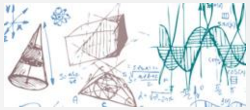


Statistics and Data Handling

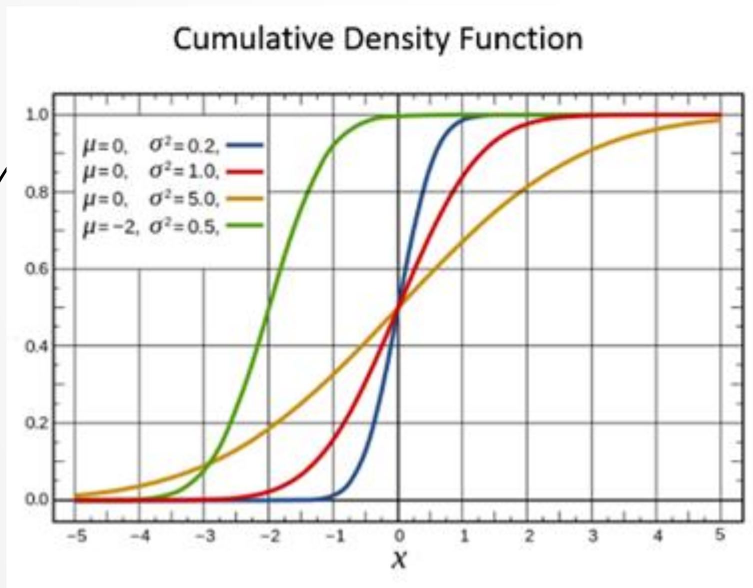
Tomasz Szumlak, Agnieszka Obłąkowska-Mucha

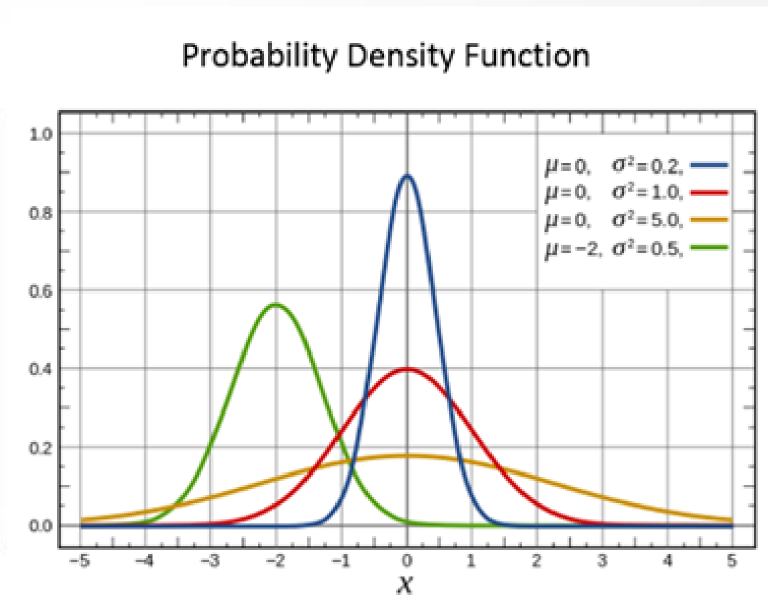
WFiIS AGH
Kraków



Before we go further…

- Recap and reminder
- In statistics we deal with two major players: probability density function $\boldsymbol{PDF}$ and cumulative density function $\boldsymbol{CDF}$





More variables

- Especially important for the inference is operating with samples of measurements (data), and usually we have $\boldsymbol{n}$ of them
- Define the CDF for this case:

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$

- And the PDF in this case:

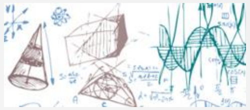
$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(x_1, x_2, \dots, x_n)$$

- Any marginal PDF of RV x_k

$$g_k(x_k) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_{k-1} dx_{k+1} \cdots dx_n$$

- … and the mean value for x_k

$$E[x_k] = \mu_k = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_k f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n$$



More variables

- And the same using the marginal PDF of x_k

$$E[x_k] = \mu_k = \int_{-\infty}^{+\infty} x_k g_k(x_k) dx_k$$

NICE!

- Now, in this convention let's write out the mean, variance and covariance

$$E[x_i] = \boldsymbol{\mu_i}$$

$$E[(x_i - E[x_i])^2] = E[(x_i - \mu_i)^2] = \boldsymbol{\sigma_i^2}$$

$$Cov(x_i, x_j) = E\left[(x_i - E[x_i])\left(x_j - E\left[x_j\right]\right)\right] = E\left[(x_i - \mu_i)\left(x_j - \mu_j\right)\right] = \boldsymbol{c_{ij}}$$

- We can also introduce a pseudo-vector notation

$$\vec{x} = \{x_1, x_2, \dots, x_n\}, \vec{X} = \{X_1, X_2, \dots, X_n\}$$

$$f(\vec{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(\vec{x})$$

Nice and compact!

More variables

- We can also put all our variances and covariances in one structure that we call **covariance matrix**

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix} \quad c_{ii} = \sigma_i^2, c_{ij} = c_{ji}$$

- Also, we can do similar thing („vectorisation”) for the means

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

$$c_{ij} = E\left[(x_i - \mu_i)(x_j - \mu_j)\right]$$

$$\boldsymbol{C} = \boldsymbol{E}[(\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}})(\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}})^{\boldsymbol{T}}]$$

- The respective elements can be written explicitly (take 2 RV)

$$\vec{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \vec{x}^T = (x_1, x_2) \quad \vec{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \vec{\mu}^T = (\mu_1, \mu_2)$$

More variables

- Now make the complete calculations

$$(\vec{x}-\vec{\mu})^T = (x_1-\mu_1, x_2-\mu_2), \vec{x}-\vec{\mu} = \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$E[(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T] = \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} (x_1-\mu_1, x_2-\mu_2) =$$

$$= \begin{pmatrix} (x_1-\mu_1)(x_1-\mu_1) & (x_1-\mu_1)(x_2-\mu_2) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)(x_2-\mu_2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & c_{12} \\ c_{21} & \sigma_2^2 \end{pmatrix}$$

- We can use our new and compact notation to derive one super important rule in statistics: **error propagation formula**
- It combines variable change and multivariate functions of RV
- Interested already? Go to the next page!

Error propagation rule

- First some background… Imagine the following problem: in order to measure a quantity $\boldsymbol{y}$ (can be more than one of these) we measure $\boldsymbol{n}$ R.V.s $\boldsymbol{x_i}$:
  $$\vec{x} = (x_1, x_2, \ldots, x_n)$$
- So, we can define a joint PDF, $f(\vec{x})$, but again usually we do not know its form but instead we can estimate respective mean values $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)$ and covariance matrix c_{ij}
- Ok, back to our $y(\vec{x})$. In principle we could follow the whole procedure of the variable change, but we can also just live with estimating the mean $\boldsymbol{E}[\boldsymbol{y}]$ and variance $\boldsymbol{V}[\boldsymbol{y}]$.
- The technique to be applied relies on using Taylor series expansion about the mean values of $\vec{x}$ (like asking what are the typical y value for typical x_i?)

$$y(\vec{x}) = \sum_{k/0}^{k/\infty} \frac{y^{(k)}(\vec{x}=\vec{c})}{k!}(\vec{x}-\vec{c})^k$$

$$y(\vec{x}) = y(\vec{\mu}) + \left(\frac{\partial y}{\partial x_1}\right)_{x_1=\mu_1} (x_1 - \mu_1) + \cdots + \left(\frac{\partial y}{\partial x_n}\right)_{x_n=\mu_n} (x_n - \mu_n) + \cdots$$

Error propagation rule

- We stop the expansion after the first element

$$y(\vec{x}) \approx y(\vec{\mu}) + \sum_{l/1}^{l/n} \left(\frac{\partial y}{\partial x_l}\right)_{x_l=\mu_l} (x_l - \mu_l)$$

- In principle we could expand it about any point, but we use the fact that $E[(x_l - \mu_l)] = 0$ (**sneaky!**), $\boldsymbol{E[y] \approx y(\vec{\mu})}$
- $\sigma_y^2 = E[y^2] - E^2[y]$, so… we just need to know the first term

$$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) + 2y(\vec{\mu}) \cdot \sum_{l/1}^{l/n} \left(\frac{\partial y}{\partial x_l}\right)_{x_l=\mu_l} E[(x_l - \mu_l)] +$$

$$+E\left[\left(\sum_{l/1}^{l/n} \left(\frac{\partial y}{\partial x_l}\right)_{x_l=\mu_l} (x_l - \mu_l)\right)\left(\sum_{j/1}^{j/n} \left(\frac{\partial y}{\partial x_j}\right)_{x_j=\mu_j} (x_j - \mu_j)\right)\right]$$

These are just numbers!

$$\sum_{l,j/1}^{l,j/n} \left(\frac{\partial y}{\partial x_l}\right)\left(\frac{\partial y}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} E\left[(x_l - \mu_l)(x_j - \mu_j)\right]$$

Error propagation rule

- Finally, we get

$$E[y^2(\vec{x})] \approx y^2(\vec{\mu}) + \sum_{l,j/1}^{l,j/n} \left(\frac{\partial y}{\partial x_l}\right)\left(\frac{\partial y}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} c_{lj}$$

Covariance matrix

$$\sigma_y^2 \approx \sum_{l,j/1}^{l,j/n} \left(\frac{\partial y}{\partial x_l}\right)\left(\frac{\partial y}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} c_{lj}$$

This is related to all $\boldsymbol{x_l}$

- And beyond… We can have many composite variables that depend on what we measure: $\vec{y} = \big(y_1(\vec{x}), y_2(\vec{x}), \dots, y_m(\vec{x})\big)$, so we can get a covariance matrix for all $\boldsymbol{y}$s:

$$u_{kl} \approx \sum_{i,j/1}^{i,j/n} \left(\frac{\partial y_k}{\partial x_i}\right)\left(\frac{\partial y_l}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} c_{ij}$$

That's it! We propagate $\boldsymbol{x}$ erros to get $\boldsymbol{y}$ erros

Note the change of indices!

Error propagation rule

- And an elegant matrix form

$$\mathcal{U} = \mathcal{T}\mathcal{C}\mathcal{T}^T \qquad t_{ij} = \left(\frac{\partial y_i}{\partial x_j}\right)_{\vec{x}=\vec{\mu}}$$

$\mathcal{U}$: **y** uncertainty

$\mathcal{C}$: **x** uncertainty

$\mathcal{T}$: Transformation from $\boldsymbol{x} \rightarrow \boldsymbol{y}$

- Very often we deal with $\boldsymbol{n}$ measurements – which can be treated as independent R.V. (I.R.V.), the consequence is that all terms off the diagonal in the covariance matrix are 0, or we have some function that depend on $\boldsymbol{n}$ I.R.V. ($c_{ii} = \sigma_i^2, c_{ij} = 0$)

$$\sigma_y^2 \approx \sum_{j/1}^{j/n} \left(\frac{\partial y}{\partial x_j}\right)^2_{\vec{x}=\vec{\mu}} \sigma_j^2$$

$$u_{kl} \approx \sum_{j/1}^{j/n} \left(\frac{\partial y_k}{\partial x_j}\right)\left(\frac{\partial y_l}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} \sigma_j^2$$

Using the rule

- Let's assume a very simple example: $y(\vec{x}) = x_1 + x_2$ and $y(\vec{x}) = x_1 \cdot x_2$, by applying the rule directly we have:

$$\sigma_{y+}^2 = \sum_{l,j/1}^{l,j/n} \left(\frac{\partial y}{\partial x_l}\right)\left(\frac{\partial y}{\partial x_j}\right)_{\vec{x}=\vec{\mu}} c_{lj} = \sigma_1^2 + \sigma_2^2 + \mathbf{2c_{12}}$$

$$\sigma_{y\cdot}^2 = y^2\left(\frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} + \frac{\mathbf{2c_{12}}}{x_1 x_2}\right)$$

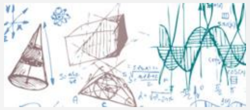
- The covariance is sensitive to addition/subtraction **if RV are not independent!**
- So, we have a very nice tool to handle our data which will help us to calculate uncertainties. If the form of the transformations or the functions are not well approximated by linear formulas then our assumptions break and we should use the confidence interval instead (see future lectures!)

Something for ML Enthusiasts

- We saw, that when propagated errors the crucial role is played by the transformations (functions)
- One simple linear transformation is rotation in 2d space which is very popular in data analysis, computer vision etc.
- Formally we call it an orthogonal transformation
- When using ML to solve problems we find that we are having way too many variables – it would be useful to reduce them!
- One way to do that is **decorrelation**! This, as we see can be interpreted as just rotation.
- The task: we have n RVs $(x_1, x_2, \ldots, x_n)$ and the covariance matrix has off-diagonal elements that are not all equal 0, we want to find a new set of RVs $(y_1, y_2, \ldots, y_n)$ for which $u_{ij} = 0$
- We postulate it is always possible with a linear transformation like this

$$y_i = \sum_{j/1}^{j/n} t_{ij} x_j$$

Something for ML Enthusiasts



- Let's calculate the covariances for ys

Pay attention to the indices!

$$u_{ij} = cov[y_i, y_j] = cov\left[\sum_{l/1}^{l/n} t_{\boldsymbol{il}} x_l , \sum_{k/1}^{k/n} t_{\boldsymbol{jk}} x_k\right] =$$

$$= \sum_{l,k/1}^{l,k/n} t_{il} t_{jk} cov[x_l, x_k] = \sum_{l,k/1}^{l,k/n} t_{il} c_{lk} \boldsymbol{t^T_{kj}}$$

When tanspose: change the order of multiplication and inverse the indices!

- Ok guys… **we are back at the error propagation formula**!
- Our task then is to find a matrix $\mathcal{T}$ to make $\mathcal{U} = \mathcal{T}\mathcal{C}\mathcal{T}^T$ diagonal
- Very well known problem: first we need to find the e-vectors $\vec{\lambda}^{(i)}, i = 1,2,\dots,n$ of the covariance matrix $\mathcal{C}$

$$\mathcal{C}\vec{\lambda}^{(i)} = \lambda_i \vec{\lambda}^{(i)}$$

- In this procedure the e-vectors are determined up to a multiplicative factor, which can be set by requiring all $\vec{\lambda}^{(i)}$ should have unit length

Something for ML Enthusiasts

- When a matrix is symmetric the e-vecs are always orthogonal
- This is always true for the covariance matrix! So, we have

$$\vec{\lambda}^{(i)} \cdot \vec{\lambda}^{(j)} = \sum_{k/1}^{k/n} \lambda_k^i \lambda_k^j = \delta_{ij}$$

- We can proceed as follow: rows of the $\mathcal{T}(=\lambda_j^i)$ matrix are the e-vectors, and the columns of $\mathcal{T}^T(=\lambda_i^j)$ are the e-vectors, then

$$u_{ij} = \sum_{l,k/1}^{l,k/n} t_{il} c_{lk} t_{kj}^T = \sum_{l,k/1}^{l,k/n} \lambda_i^l \boldsymbol{c_{lk}} \lambda_j^k = \sum_{l/1}^{l/n} \lambda_i^l \boldsymbol{\lambda_j} \lambda_j^l = \lambda_j \vec{\lambda}^{(i)} \cdot \vec{\lambda}^{(j)}$$

$$u_{ij} = \lambda_j \delta_{ij}$$

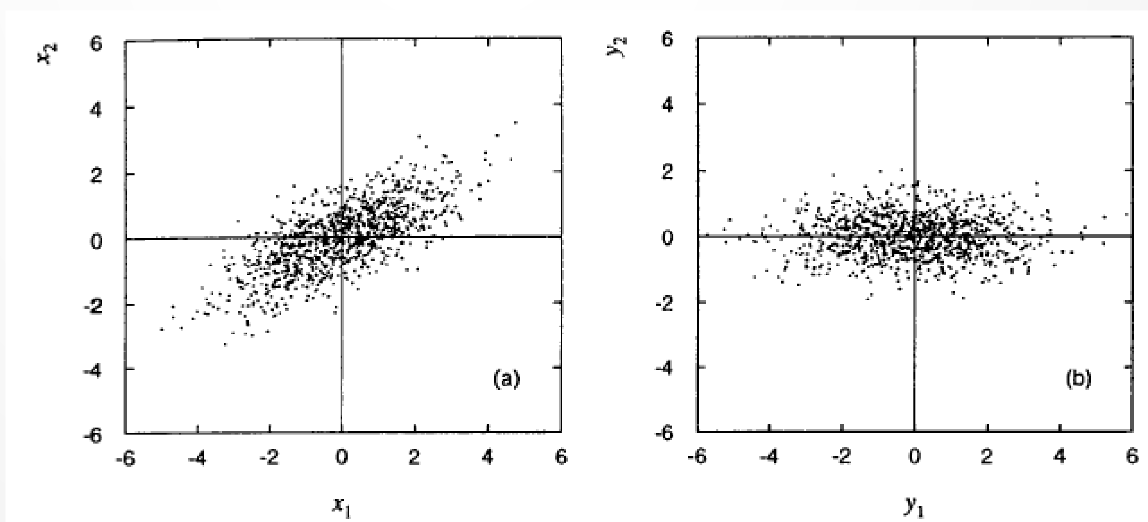
- Variances of new RVs are expressed as e-values of the original covariance matrix $\mathcal{C}$ and $\mathcal{T}\mathcal{T}^T = 1$, thus, $\mathcal{T}^T = \mathcal{T}^{-1}$

Something for ML Enthusiasts

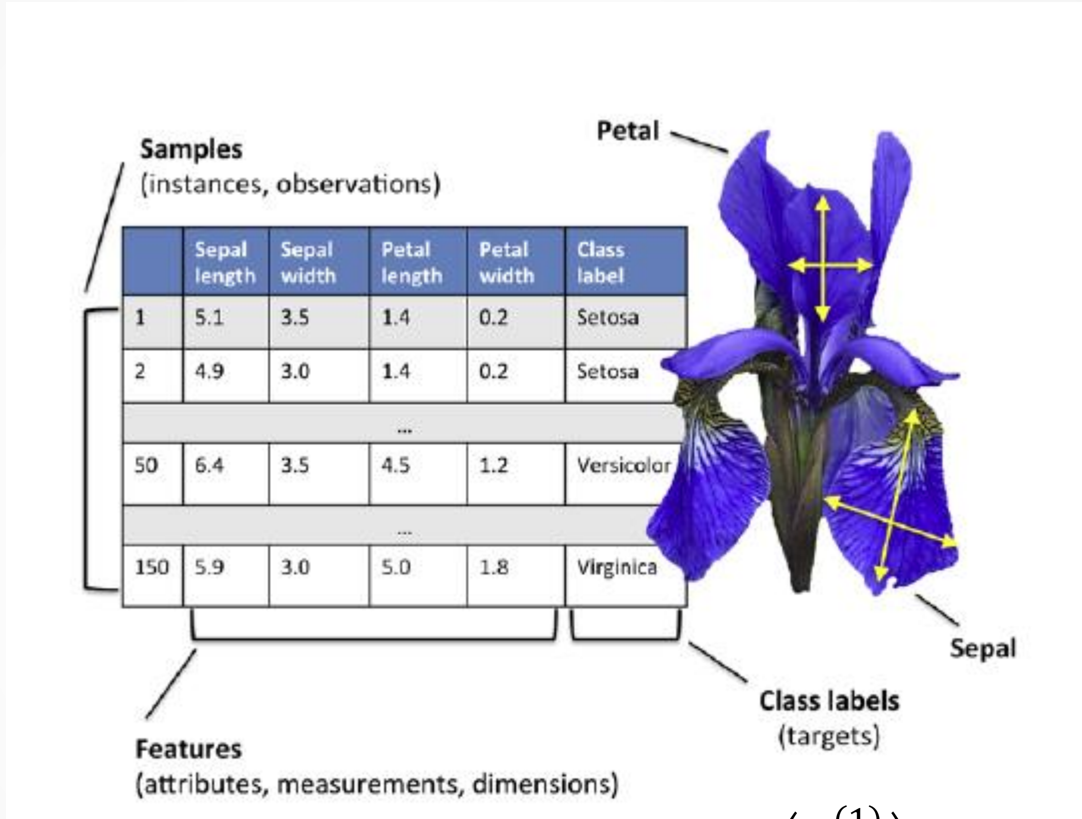
- For two dimensions this is a simple calculation, for more we just use computer programs. In the case of 2d it can be shown:

$$C = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} \quad \vec{\lambda}^{(1)} = \begin{pmatrix} cos\theta \\ sin\theta \end{pmatrix} \quad \vec{\lambda}^{(2)} = \begin{pmatrix} -sin\theta \\ cos\theta \end{pmatrix}$$

$$\theta = \frac{1}{2} tg^{-1}\left(\frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}\right)$$



Something for ML Enthusiasts



	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

$x^{(i)} = \left(x_1^{(i)}, \dots, x_n^{(i)}\right)$ - one instance $\quad x_j = \begin{pmatrix} x_j^{(1)} \\ . \\ . \\ . \\ x_j^{(m)} \end{pmatrix}$

Something for ML Enthusiasts