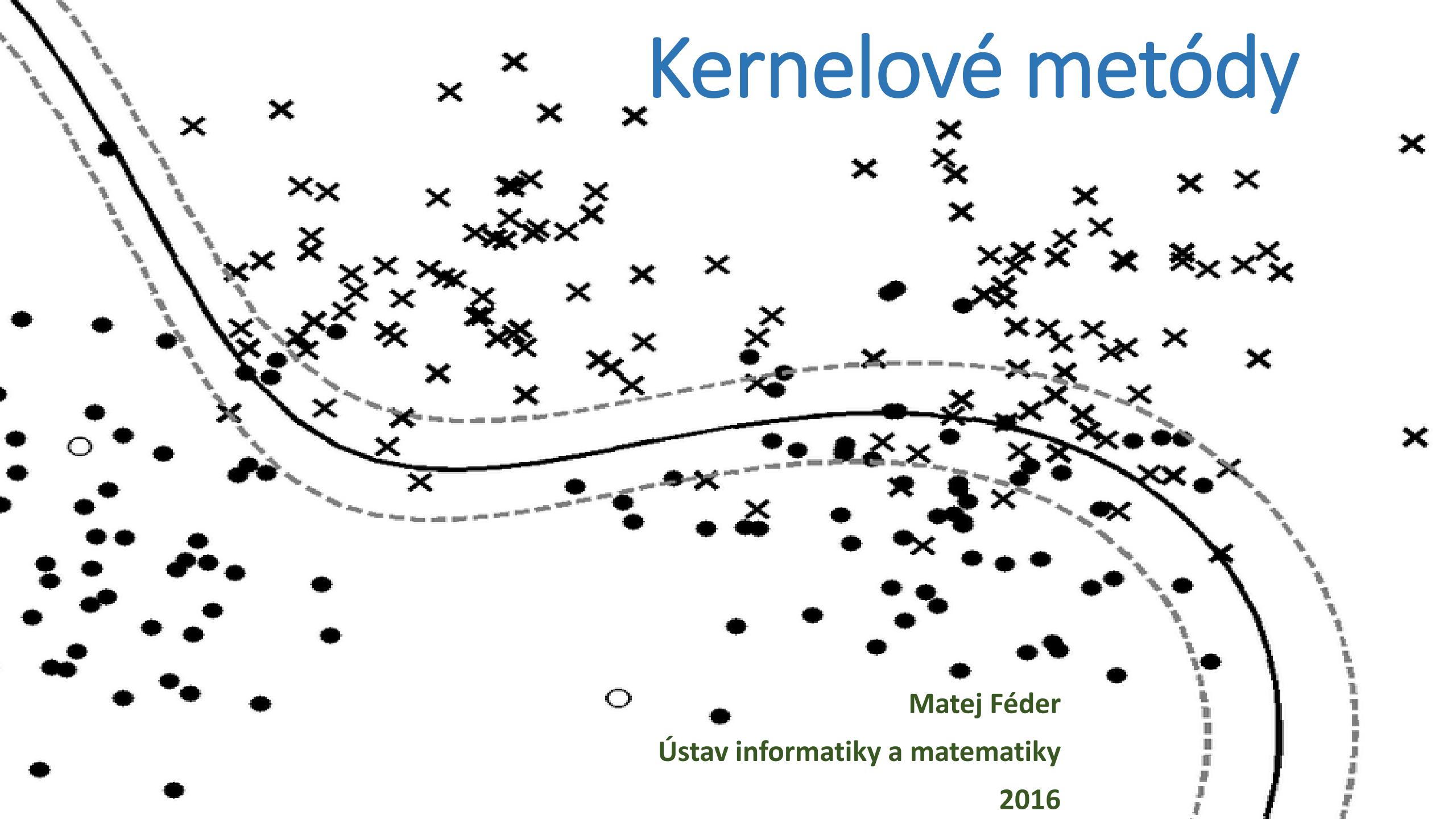


Kapitola 9

Kernelové metódy

Kernelové metódy



Matej Féder

Ústav informatiky a matematiky

2016

Obsah prednášky:

- Strojové učenie
- Kernelové metódy
 - KPCA (*PCA*)
 - GDA (*LDA*)
 - SVM
- Využitie kernelových metód

Strojové učenie - Machine Learning

Strojové učenie

Spam Detection

Credit Card Fraud Detection

Digit Recognition

Speech Understan./Recognition

Face Detection/Recognition

Product Recommendation

Medical Diagnosis

Stock Trading

.....

<https://www.coursera.org/course/ml>

Kernelové metódy

Široká škála využiteľnosti
kernelových metód:

Strojové učenie

Image processing

Redukcia dimenzie dát

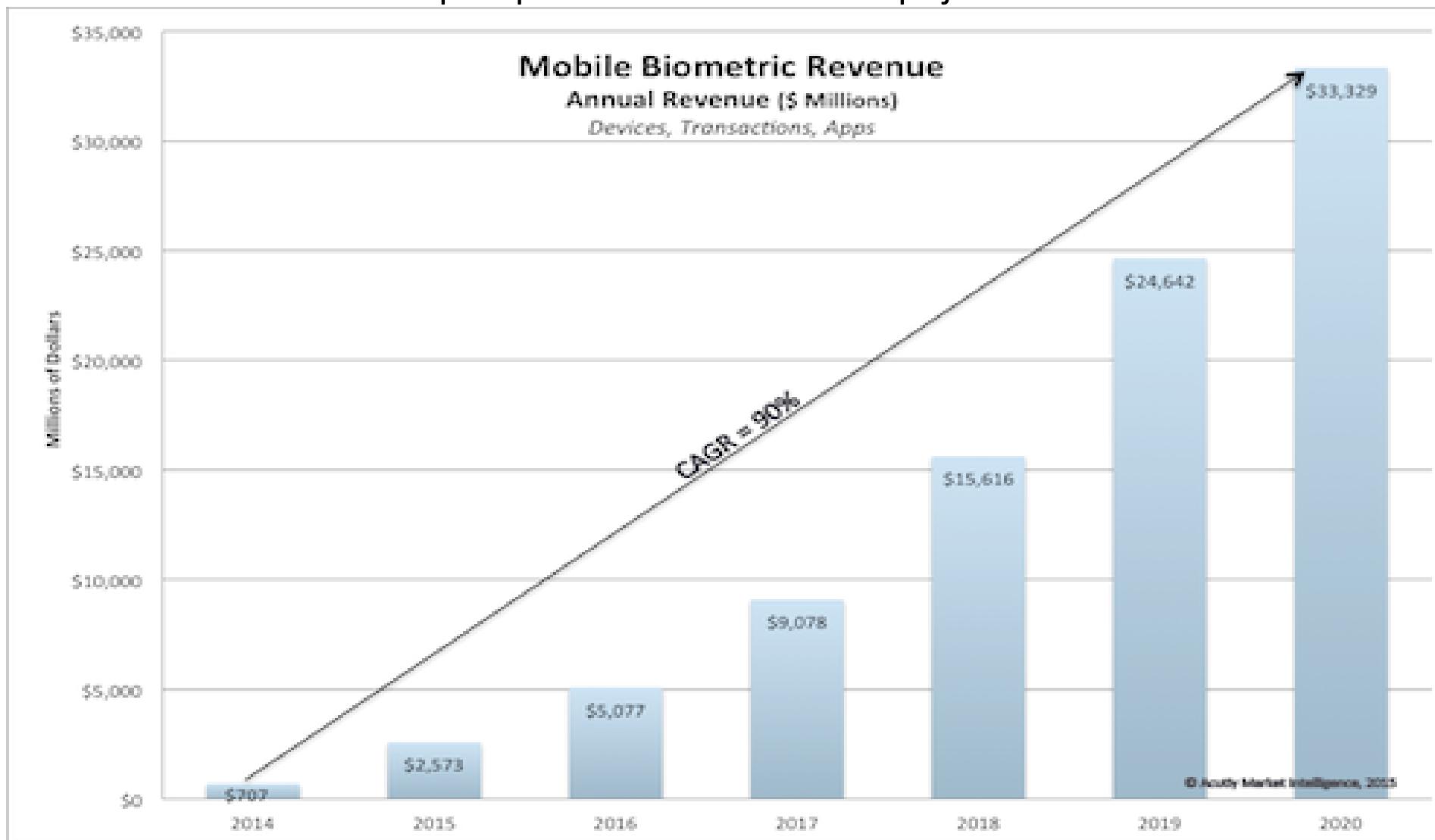
.....

<http://machinelearningmastery.com/practical-machine-learning-problems/>

\$Motivácia

Predikcia príjmov z mobilnej biometrie do roku 2020:

- 90% nárast príjmov,
- predpoklad: 33 miliárd USD príjmov z biometrie v r.2020.



<http://www.youtube.com/watch?v=ImPJeKRs8gE&feature=related>

<http://www.youtube.com/watch?v=V25qu1xpJOc&feature=relmfu>

<http://www.youtube.com/watch?v=OAO8Vhl3oUU>

https://www.youtube.com/watch?v=_VFUXFirTX4

<https://www.youtube.com/watch?v=X41lWA9Mf-Q>

https://www.youtube.com/watch?v=b_DSTuxdGDw

<https://www.youtube.com/watch?v=qv6UVQ0F44>

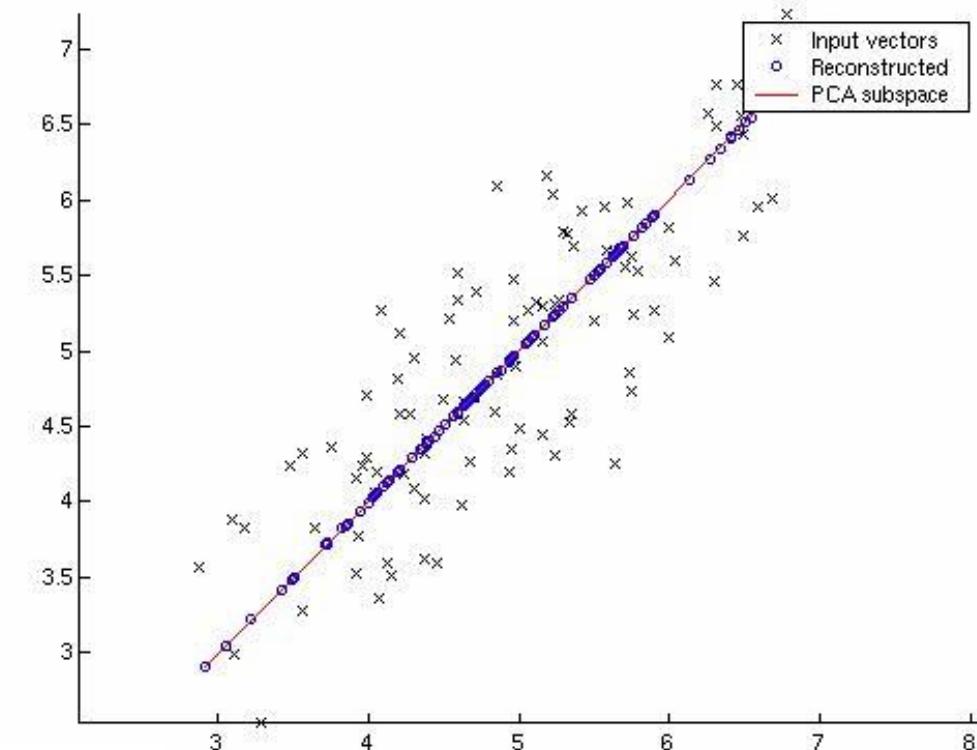
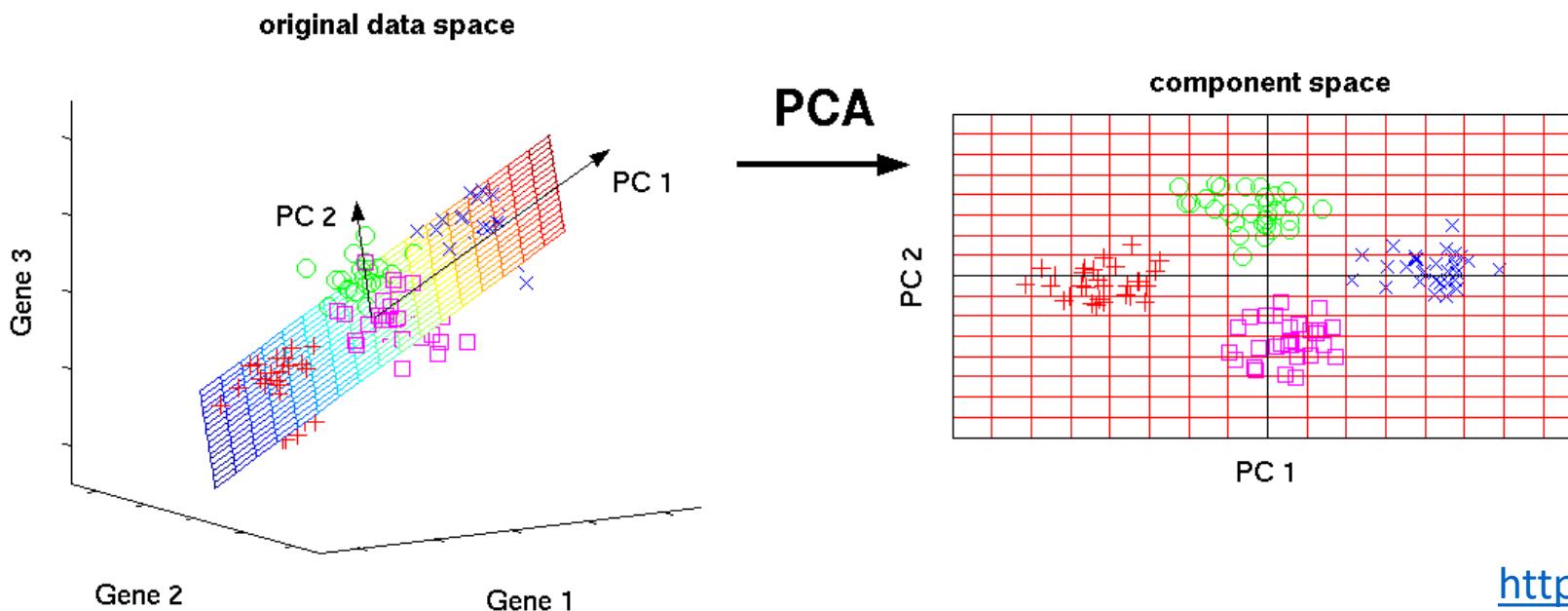
<https://www.youtube.com/watch?v=xclBoPuNliw>

10 Machine Learning based Products You MUST See:
<https://www.youtube.com/watch?v=dcZvhP-lqY4>

Principal Components Analysis (PCA)

Analýza hlavných komponentov

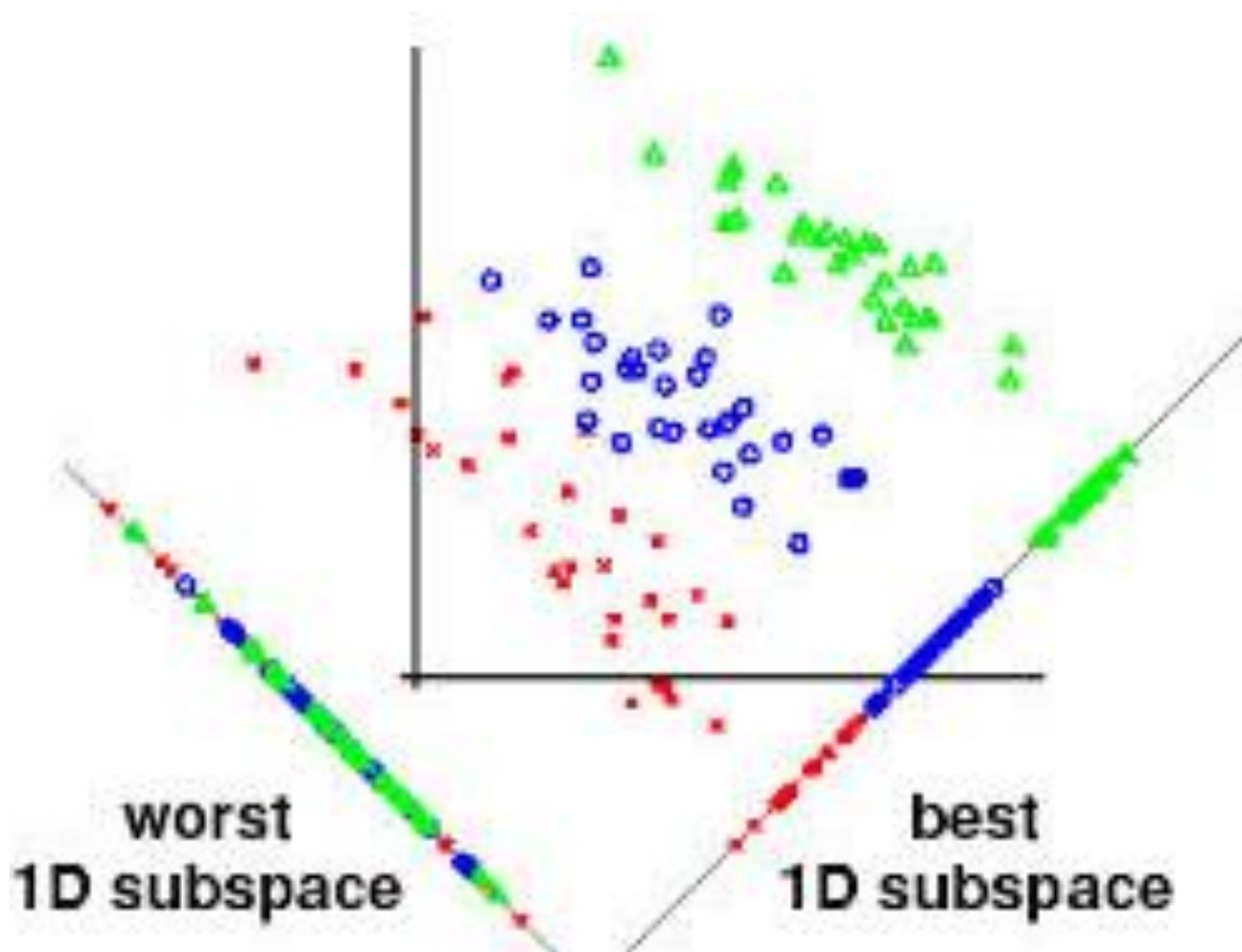
- Hotellingova transformácia
- Karhunenova-Loèvova transformácia
- široko používaná metóda v oblasti spracovania obrazu využívaná pri riešení najrôznejších úloh klasifikácie, detekcie, predikcie, extrakcie príznakov a kompresie.
- **PCA hľadá najefektívnejšie vyjadrenie dát, pričom maximalizuje ich rozptyl, a to bez ohľadu na ich triedy.**



Linear Discriminant Analysis (LDA)

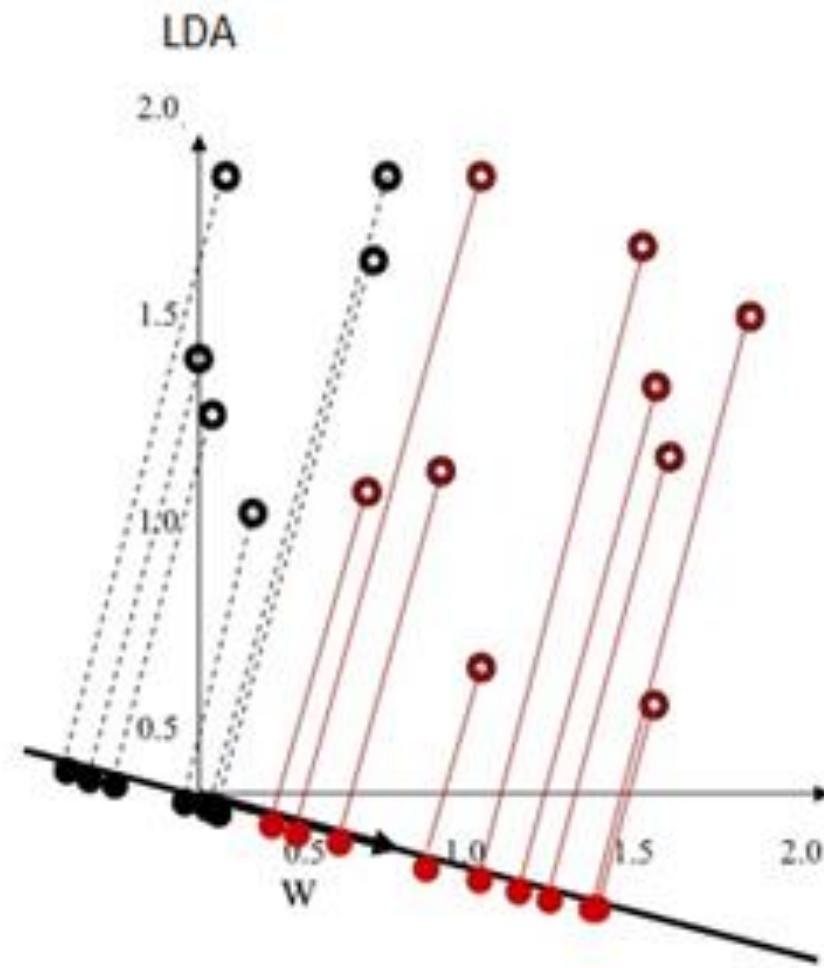
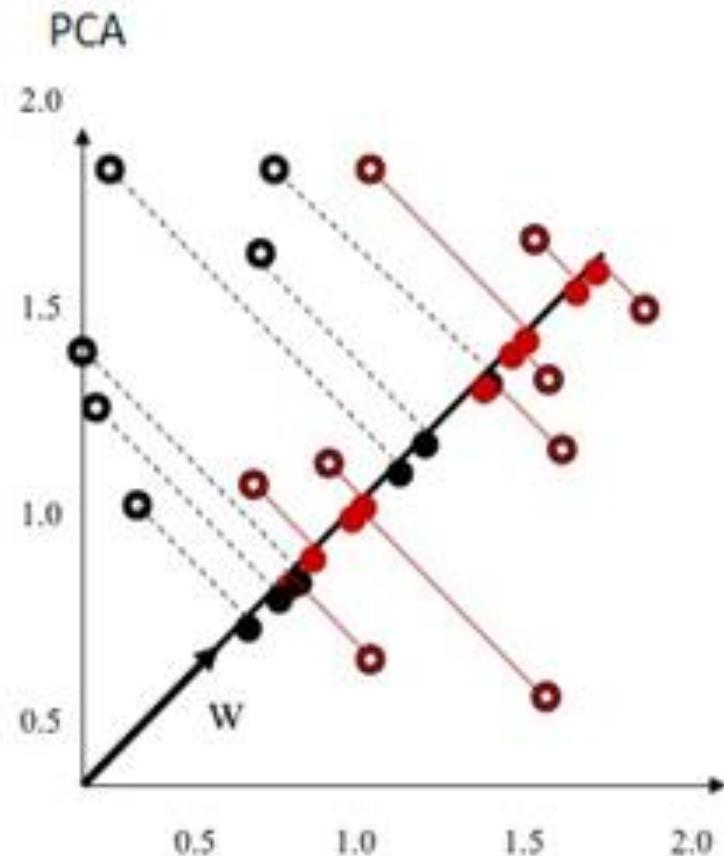
Lineárna diskriminačná analýza

- Fisherova lineárna diskriminačná analýza
 - štatistická metóda, ktorá maximalizuje medzitriedny a minimalizuje vnútroriedny rozptyl.
- toto zohľadnenie tried je veľmi výhodné v problematike rozpoznávania tvári.



PCA vs. LDA

Modelový príklad zlyhania PCA



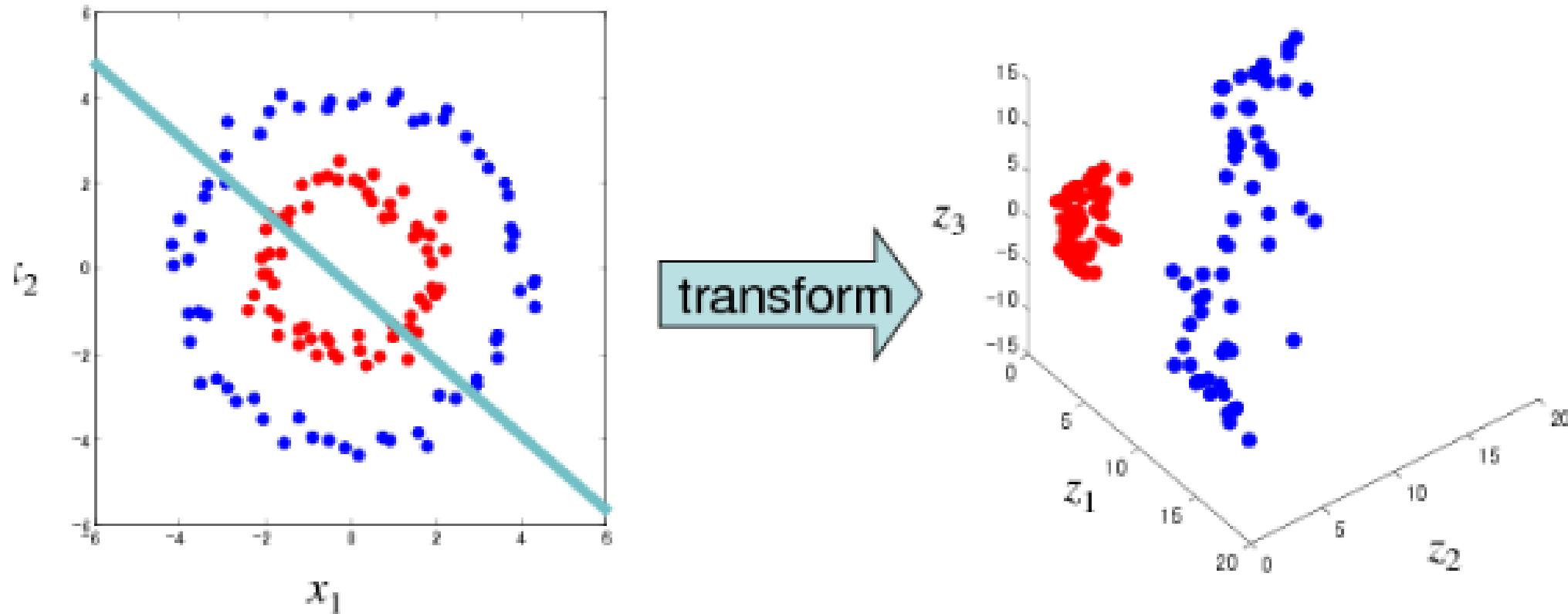
Príklad rozdelenia, premietnutia do pod priestoru bodov dvoch tried (čierne a červené krúžky) metódami PCA a LDA. PCA maximalizuje rozptyl dát bez ohľadu na príslušnosť k triede. Metóda LDA maximalizuje medzitriedny rozptyl [<http://www.csee.wvu.edu/~timm/cs591o/old/FSS.html>]

KERNELOVÉ METÓDY - MOTIVÁCIA

$$\Phi : \mathcal{R}^2 \rightarrow \mathcal{R}^3$$

Lineárne neseparovateľné

Lineárne separovateľné



$$(z_1, z_2, z_3) = (x_1^2, x_2^2, \sqrt{2}x_1x_2)$$

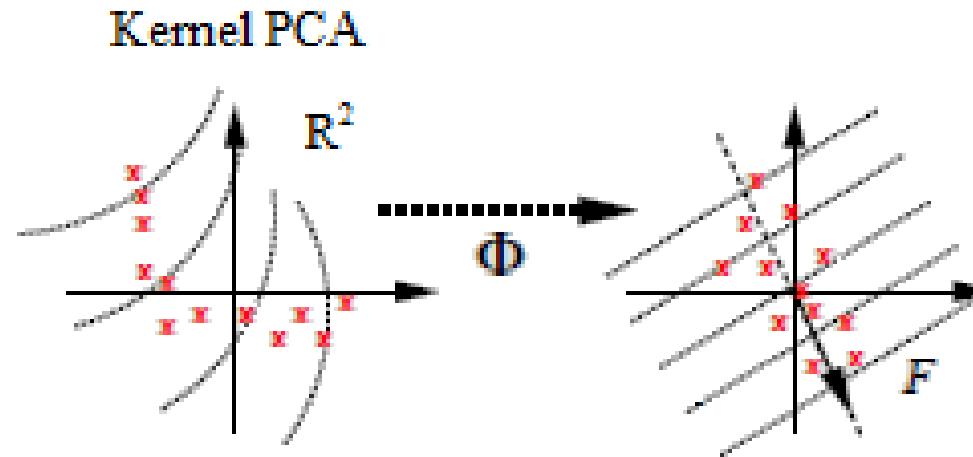
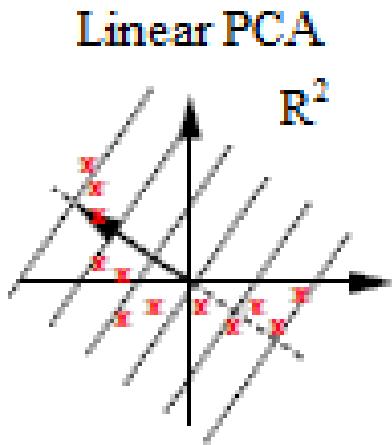
<http://www.youtube.com/watch?v=9NrALgHFwTo&feature=related>

<http://www.youtube.com/watch?v=3liCbRZPrZA>

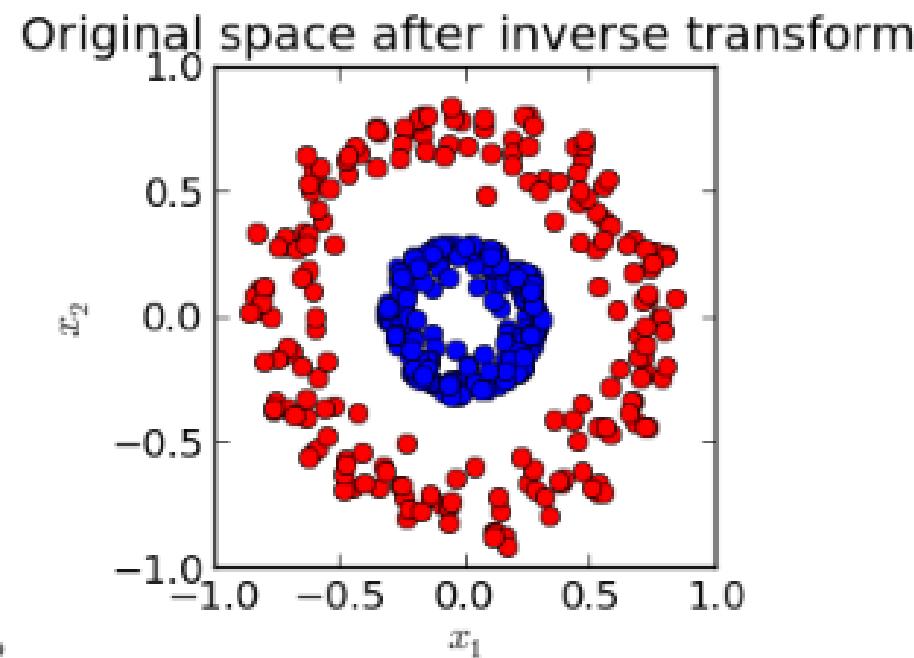
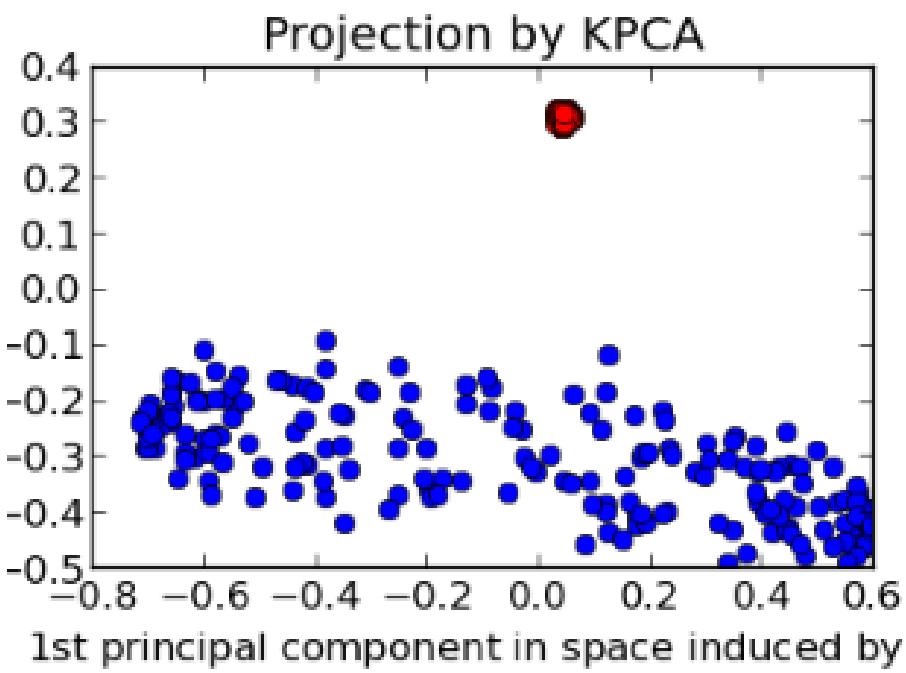
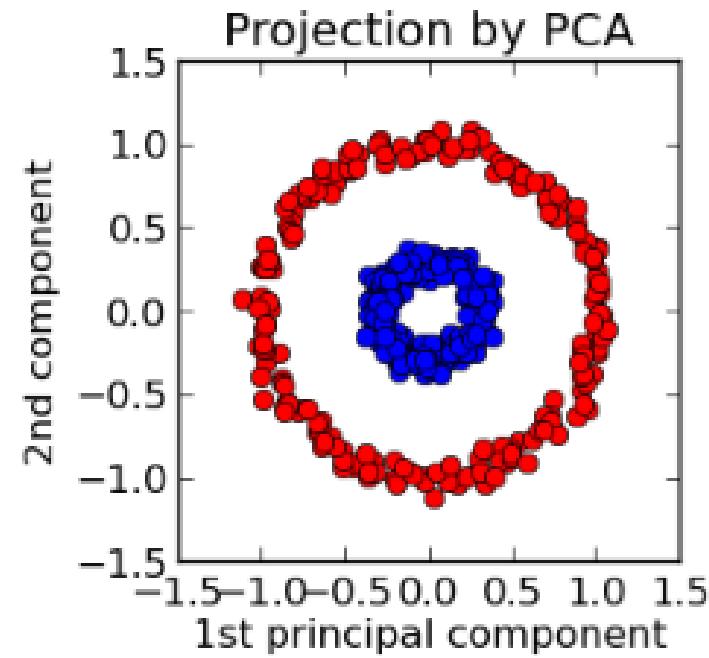
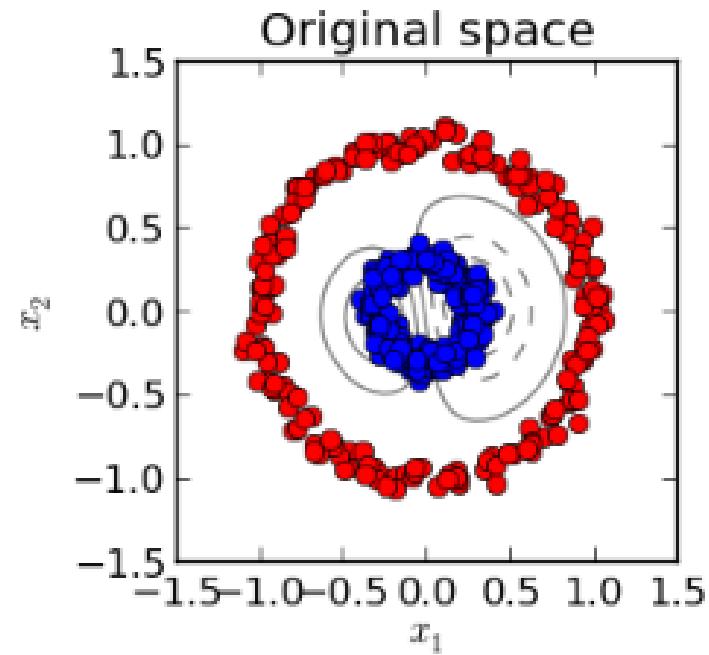
Kernelová PCA (KPCA)

Rozšírenie lineárnej PCA na KPCA

Ak je transformácia do vysokorozmerného priestoru Φ ,
potom projekciou vstupných vektorov x_1, \dots, x_n do tohto priestoru dostávame $\Phi(x_1), \dots, \Phi(x_n)$.

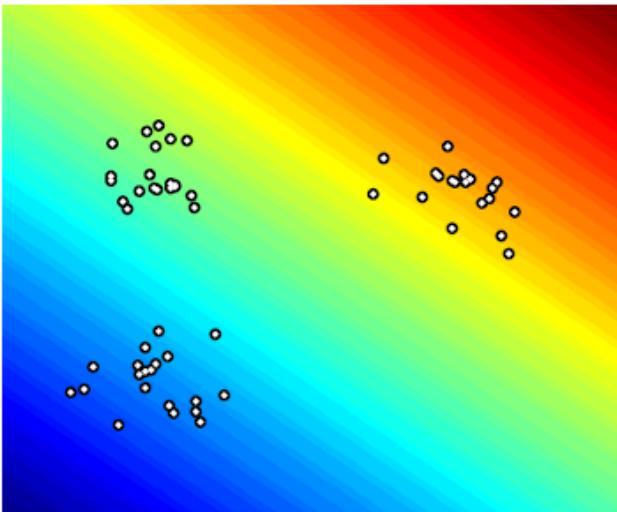


Linear PCA and the basic idea of kernel PCA: Using a suitable nonlinear mapping Φ and performing linear PCA on the mapped patterns (kernel PCA), the resulting nonlinear direction in the input space can find the most interesting direction

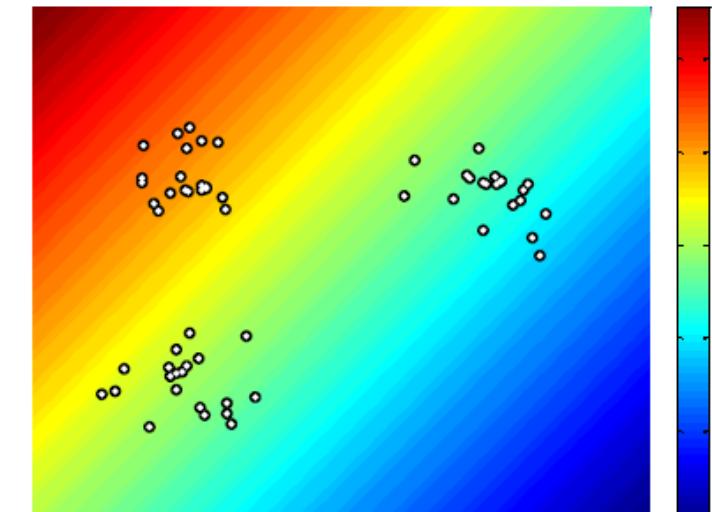


The (linear) PCA solution

PCA 1

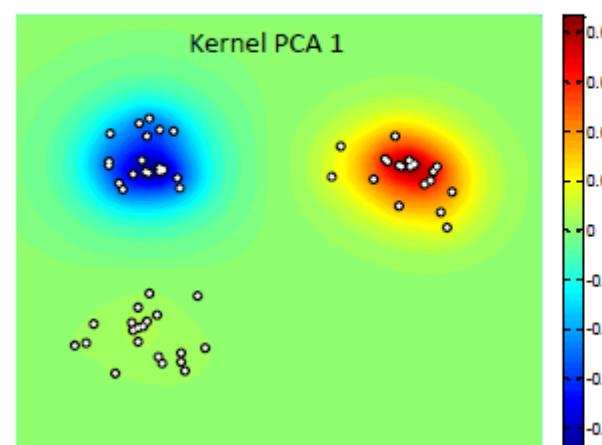


PCA 2

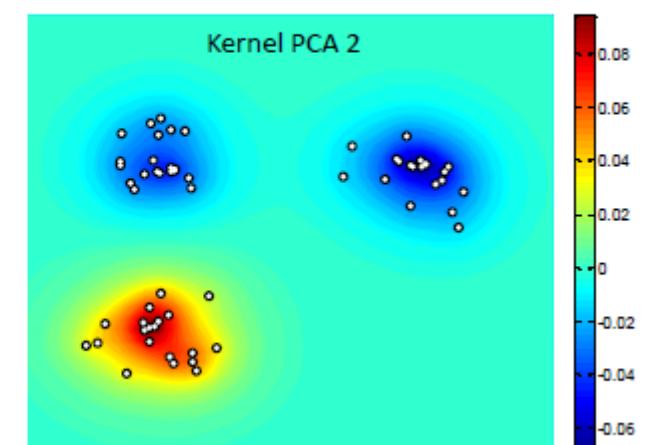


The kernel PCA solution (Gaussian Kernel)

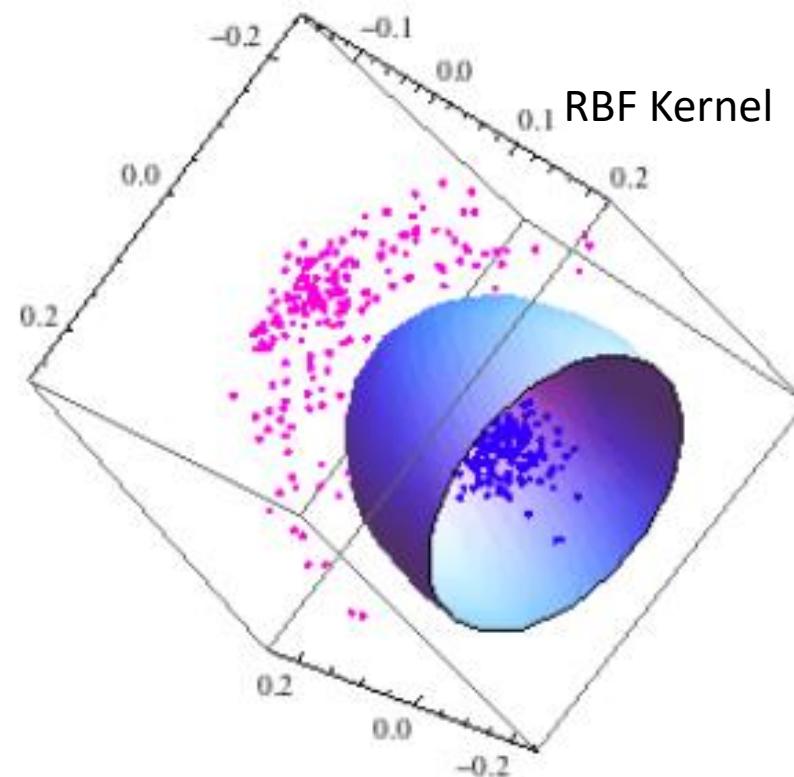
Kernel PCA 1



Kernel PCA 2



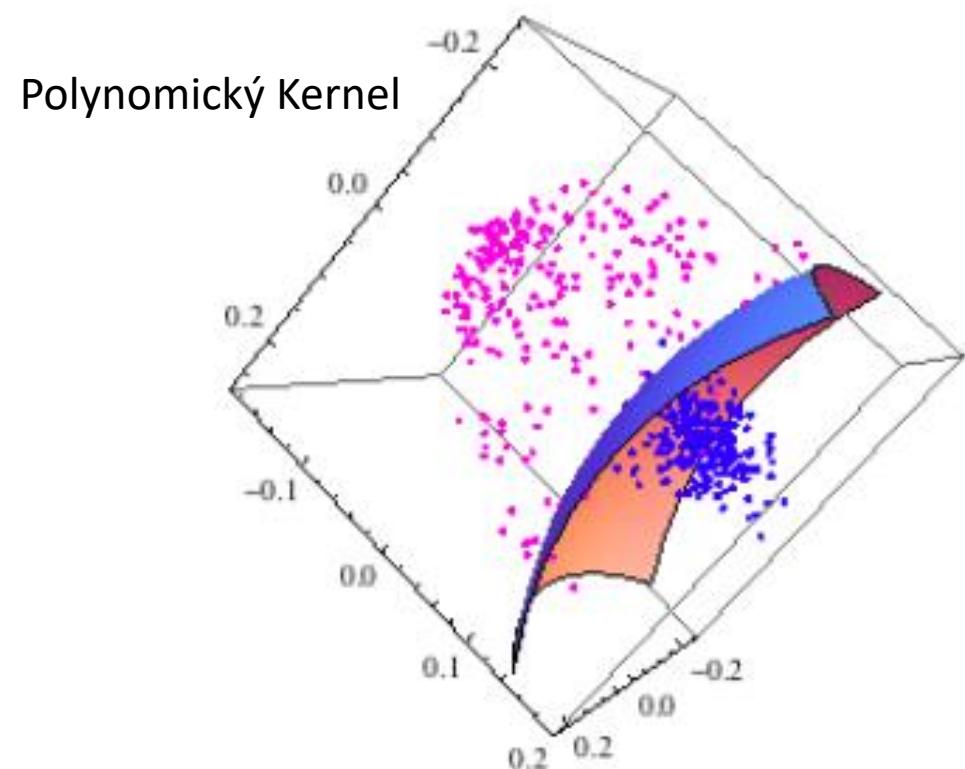
- kernelovo rozšírená LDA
- Metóda GDA si zachováva vlastnosť jej lineárneho ekvivalentu, a to maximalizáciu medzitriedneho a minimalizáciu vnútro triedneho rozptylu.
- Ďalej nadobúda vlastnosť úspešnejšej klasifikácie pre prípad zhlukových, zložitejších vstupných dát projekciou do vysokorozmerného priestoru.



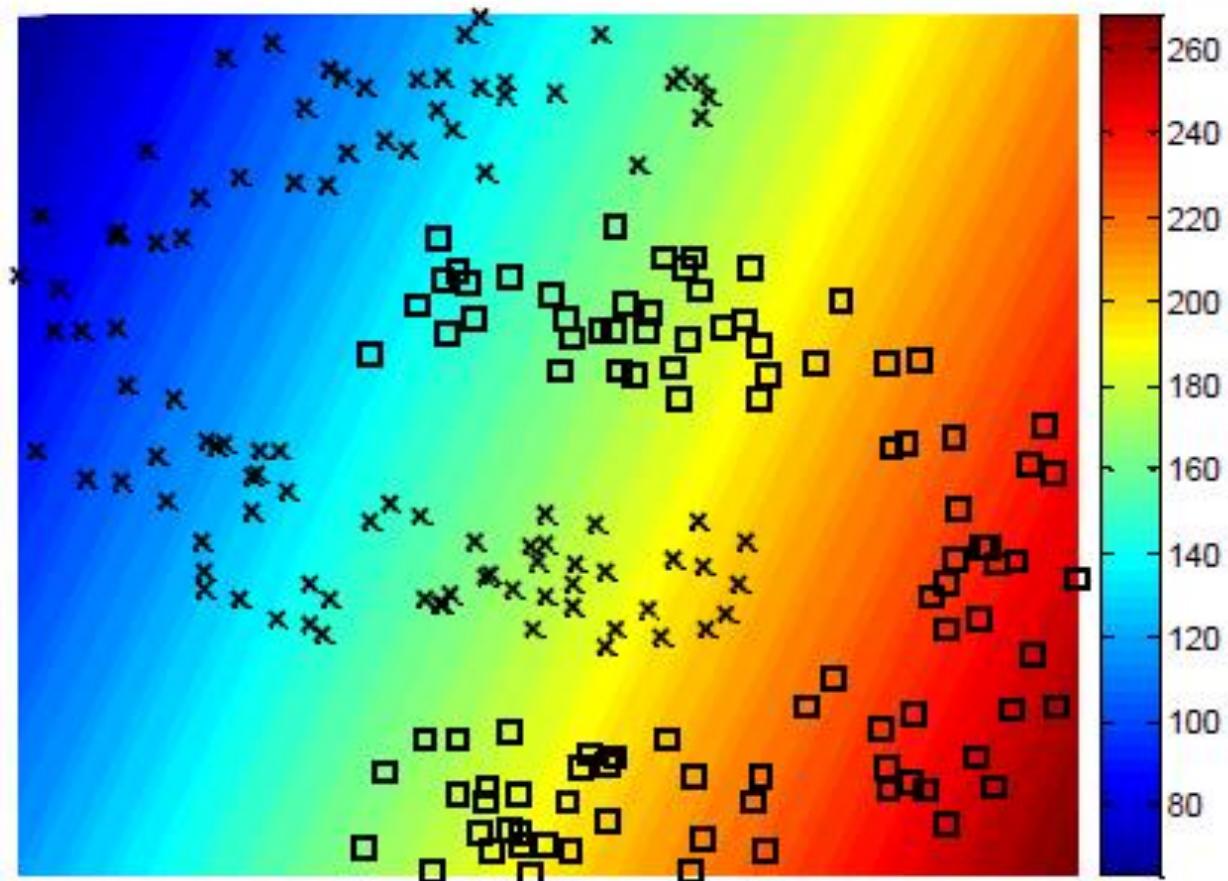
Zovšeobecnená diskriminačná analýza (GDA)

Generalized Discriminant Analysis

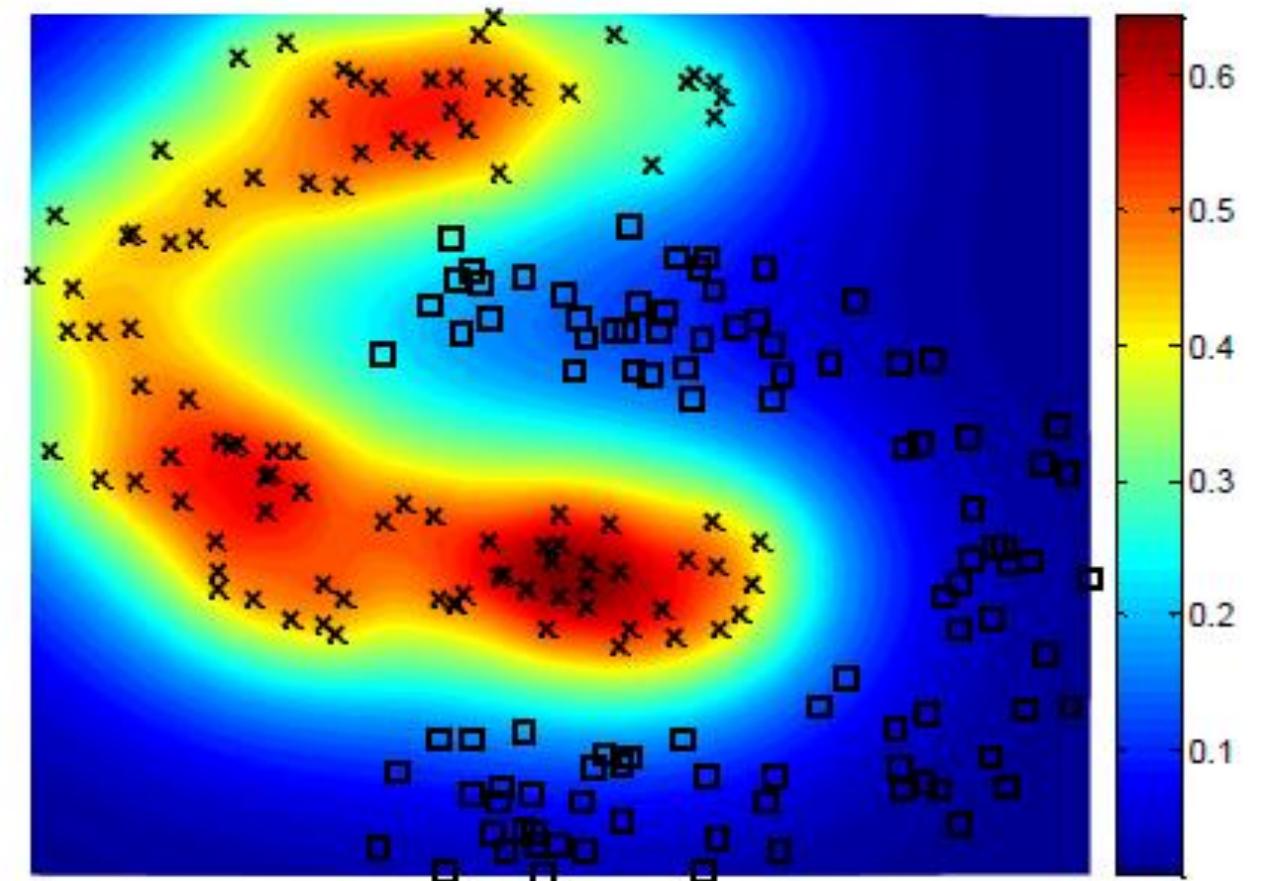
- Kernel Fisher Discriminant Analysis (KFD)



LDA



Kernel LDA



Support Vector Machine (SVM)

Stroj s podpornými vektormi



•**SVM algoritmus navrhnutý Prof. Vladimirom Vapnikom**

V. Vapnik, The Nature of Statistical Learning Theory. Springer, 1995.

•**Neskôr rozšírený o metódu „Jamného rozpetia“ (Soft margin), Vapnik a Corinna Cortes (1995)**

Corinna Cortes and V. Vapnik, "Support-Vector Networks", Machine Learning, 20, 1995.

<http://www.springerlink.com/content/k238jx04hm87j80g/>

AT&T Bell Labs

<http://www.support-vector-machines.org/>

<https://www.youtube.com/watch?v=1NxnPkZM9bc>

SVM - Support Vector Machines

- Široko využívaná metóda strojového učenia pre klasifikáciu a regresiu
- Metóda získavania optimálnej hranice medzi dvoma príkladmi jednotlivých tried

Algoritmus:

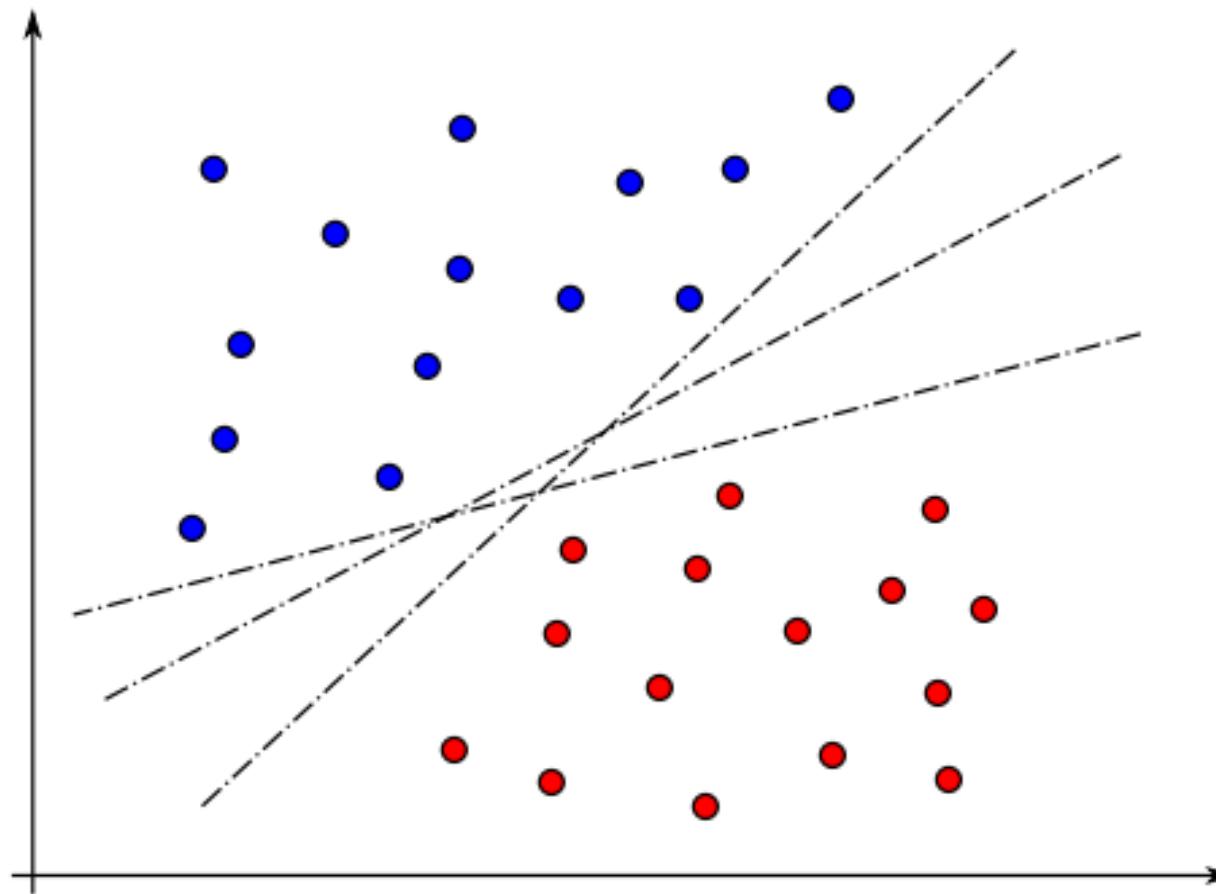
1. Definovanie optimálnej hranice, **maximalizovanie hranice medzi triedami : HARD MARGIN**

2. Rozšírenie pre lineárne neseparovateľný prípad:
pomocou penalizácie za zlú klasifikáciu : SOFT MARGIN

3. Transformácia do vysokorozmerného priestoru

SVM - Support Vector Machines

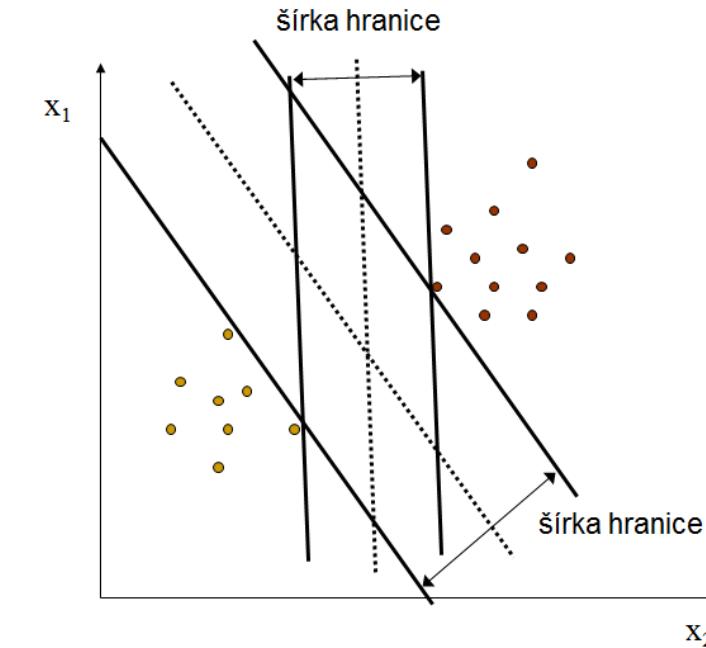
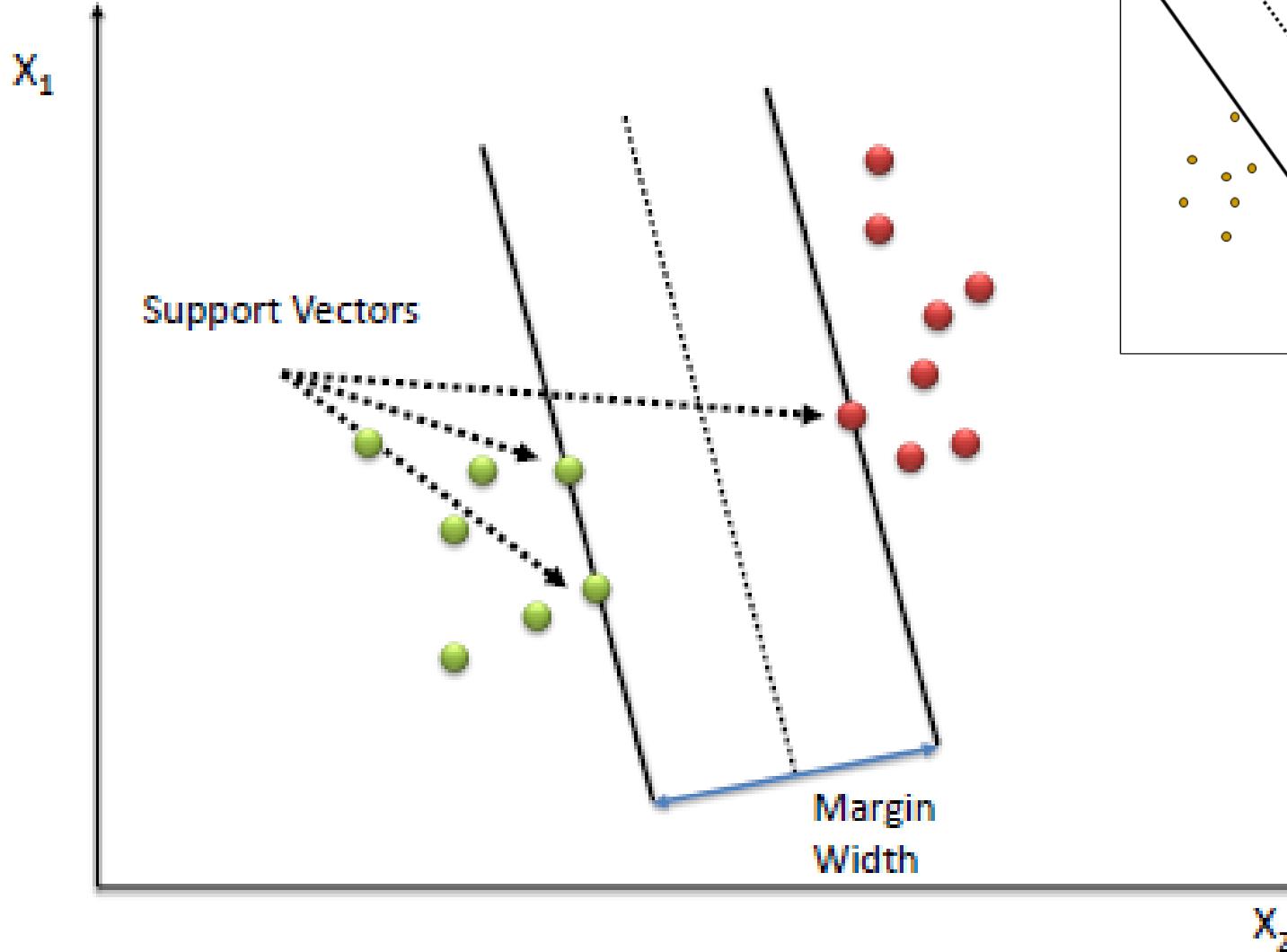
1. Definovanie optimálnej hranice, maximalizovanie hranice medzi triedami : HARD MARGIN



Lineárne separovateľný prípad :

- Treba nájsť takú rovinu ktorá zabezpečí najširšiu oddelujúcu hranicu medzi oboma triedami

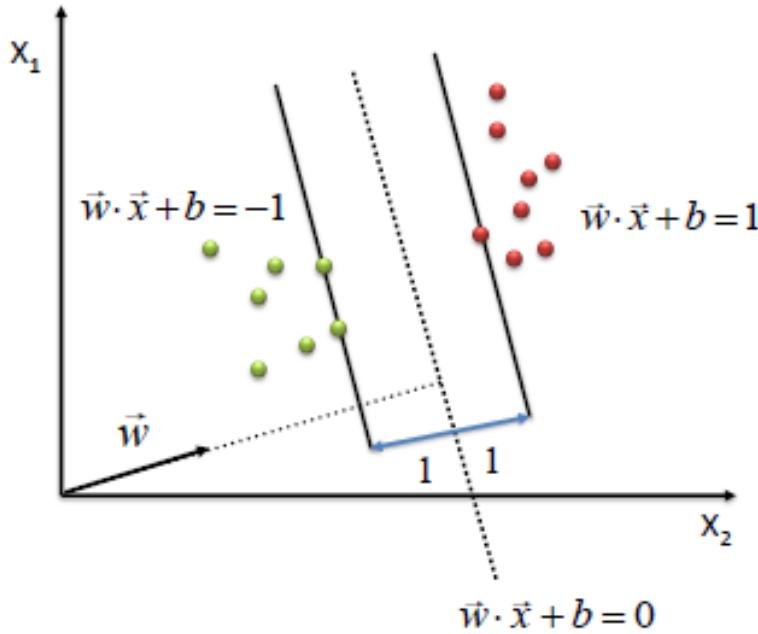
SVM - Support Vector Machines



SVM - Support Vector Machines

vektor \vec{w} je kolmý na plus-rovinu a minus-rovinu.

kde \vec{w} je vektor váhových koeficientov a b je posunutie

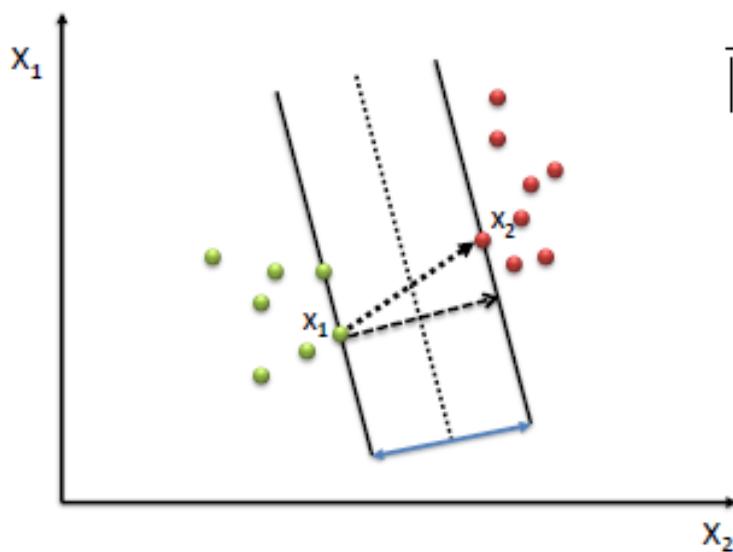


$$\boxed{\max \frac{2}{\|\vec{w}\|}}$$

s.t.

$$(w \cdot x + b) \geq 1, \forall x \text{ of class 1}$$

$$(w \cdot x + b) \leq -1, \forall x \text{ of class 2}$$



- Optimalizačný problém má tvar

$$\max \frac{2}{\|\mathbf{w}\|}$$

alebo

$$s.t. y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall \mathbf{x}_i$$

$$\min \frac{1}{2} \|\mathbf{w}\|^2$$
$$s.t. y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall \mathbf{x}_i$$

- Treba nájsť parametre \mathbf{w}, b ktoré sú riešením

$$\min \frac{1}{2} \|\mathbf{w}\|^2$$
$$s.t. y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \forall \mathbf{x}_i$$

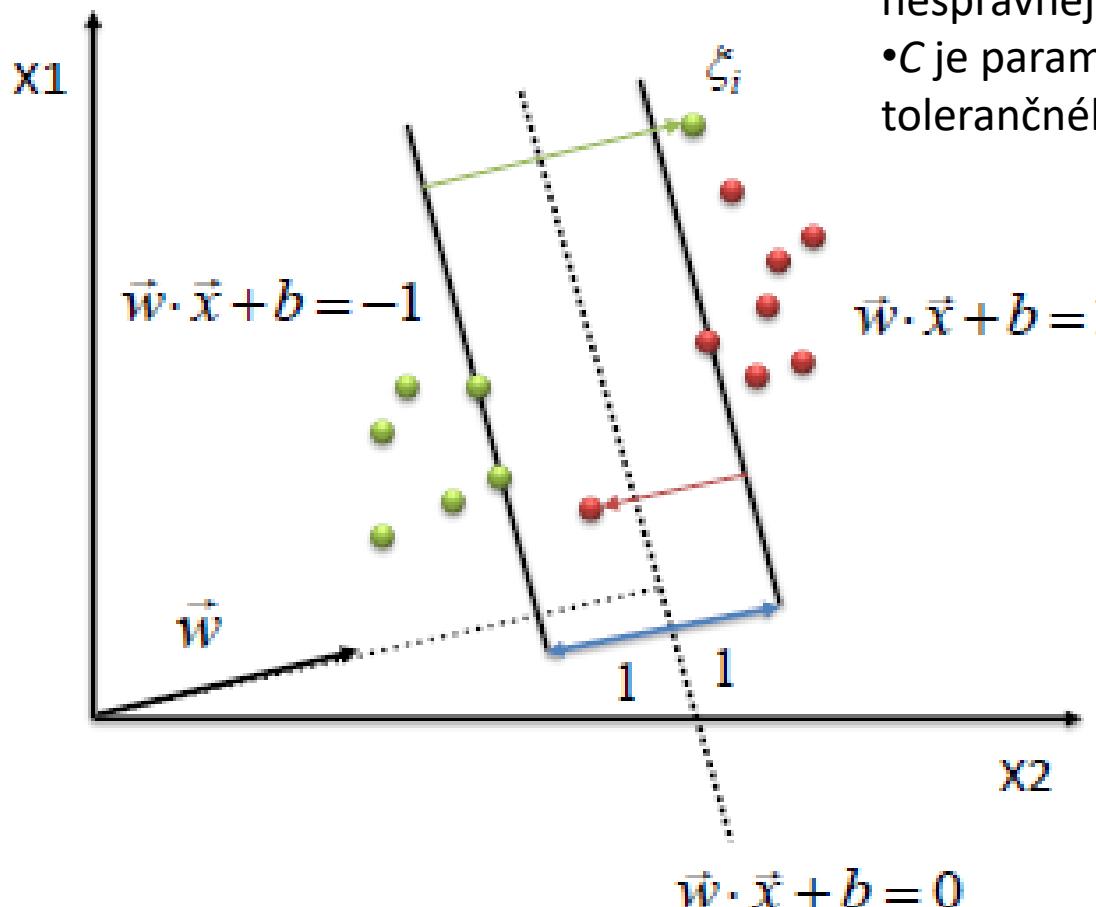
Ak sú vstupné vzory (lineárne) neseparovateľné uvedená formulácia problému nenájde riešenie

SVM - Support Vector Machines

2. Rozšírenie pre lineárne neseparovateľný prípad: pomocou penalizácie za zlú klasifikáciu :
SOFT MARGIN

kde \mathbf{w} je vektor váhových koeficientov a b je posunutie

Soft margin SVMs , L1,L2



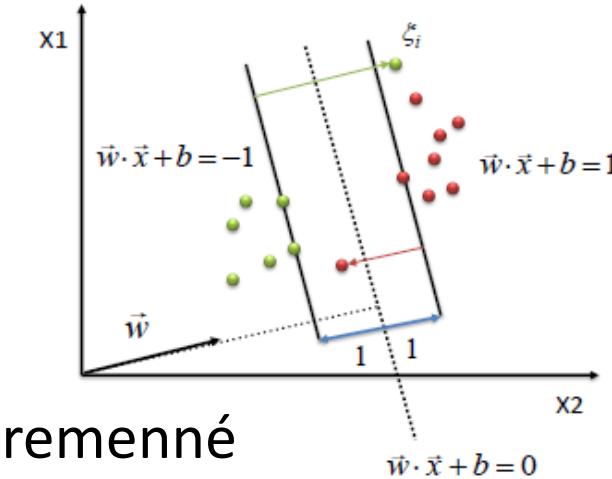
- kde KSI je relaxačný parameter (slack variable) kladné číslo, ktoré určuje mieru tolerancie nesprávnej klasikácie.
- C je parameter určujúci mieru vplyvu tolerančného pásma

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i \xi_i$$

SVM - Support Vector Machines

2. Rozšírenie pre lineárne neseparovateľný prípad: pomocou penalizácie za zlú klasifikáciu :
SOFT MARGIN

$$\min \frac{1}{2} \|\vec{w}\|^2 + C \sum_i \xi_i$$



- Algoritmus sa snaží ponechať tolerančné premenné ξ_i na nulovej hodnote pričom súčasne maximalizuje oddelujúcu hranicu
- Väčšia hodnoda $C \rightarrow \infty$, spôsobí že riešenie sa bude približovať k hard-margin SVM
- Parameter C doladuje vplyv nesprávane klasifikovaných vzorov a šírky hranice pri optimalizácii

SVM - Support Vector Machines : Hľadanie riešenia

$$\text{minimize } L_p(\mathbf{w}, b, \alpha_i) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1)$$

$$\alpha_i \geq 0$$



Lagrangián - Duálny Problém

$$\text{maximize } \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j$$

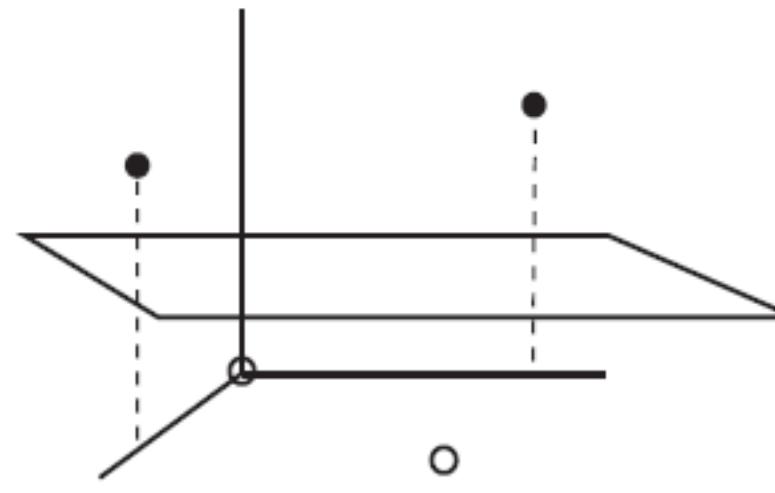
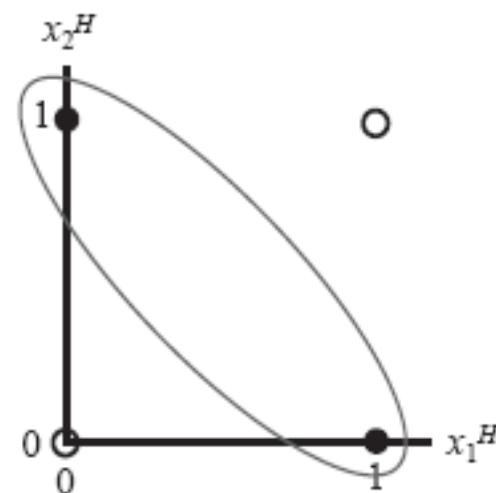
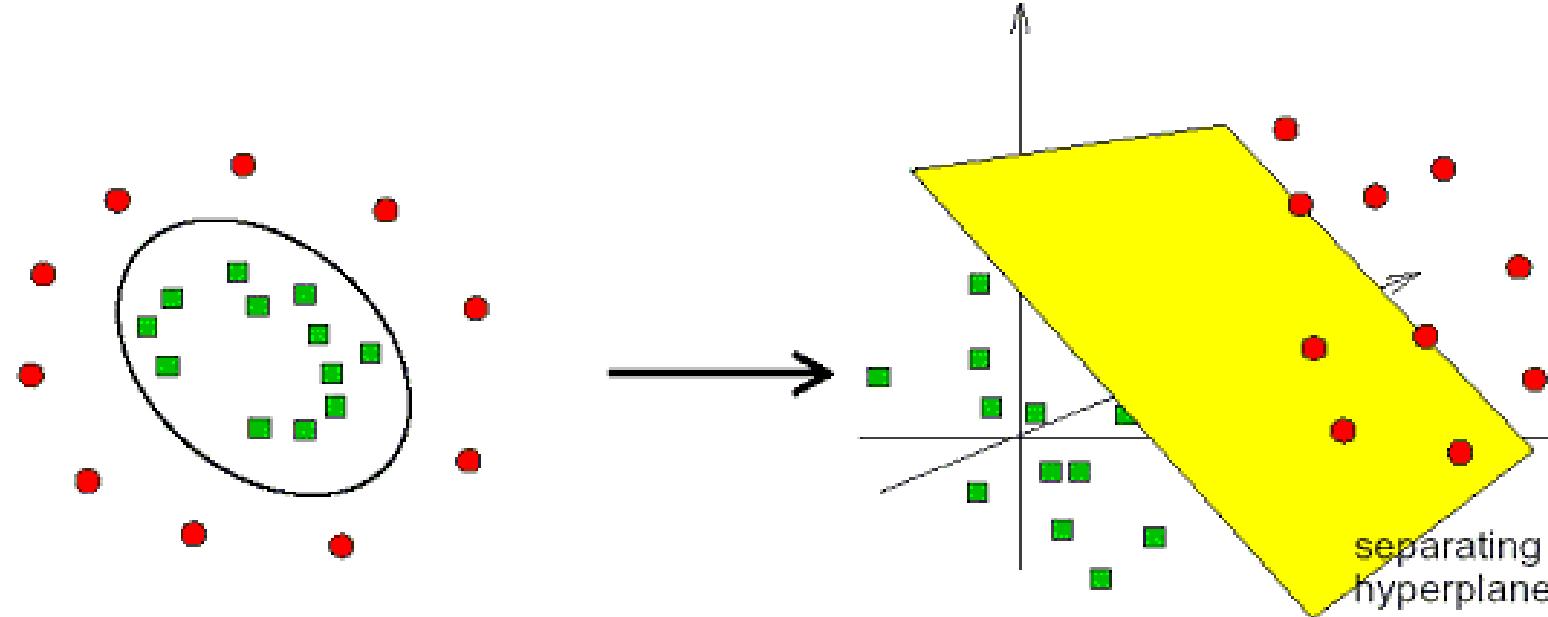
$$\alpha_i \geq 0 \text{ , and } \sum_{i=1}^n \alpha_i y_i = 0$$

- Hľadanie riešenia v sebe zahŕňa výpočet skalárneho súčinu medzi všetkými dvojicami vstupných vzorov \mathbf{x}_i \mathbf{x}_j

SVM - Support Vector Machines

3. Transformácia do vysokorozmerného priestoru

1992, [Bernhard Boser](#), [Isabelle Guyon](#) a [Vapnik](#) – nelineárna klasifikácia „Kernel trik“



SVM - Support Vector Machines : Kernel trick

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$$

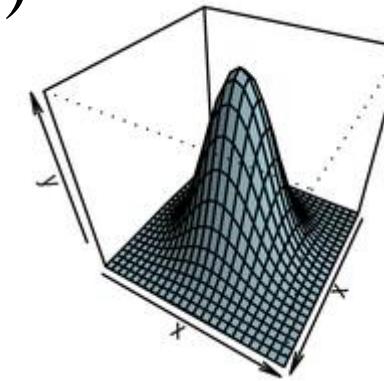
kde K je kernelová funkcia, Φ je transformácia do viacrozmerného priestoru
 $\mathbf{x}_i, \mathbf{x}_j$ sú vektory v pôvodnom priestore

□ Linearny kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$

□ Polynomiálny kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = (1 + \mathbf{x}_i^T \mathbf{x}_j)^p$

□ Gaussovský (Radial-Basis Function (RBF)) kernel:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right)$$



□ Sigmoidálny

kernel: $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\beta_0 \mathbf{x}_i^T \mathbf{x}_j + \beta_1)$

- t.j. jej výsledok predstavuje skalárny súčin obrazov v mnohorozmernom priestore, nemusíme vykonávať explicitné transformovanie vstupov do mnohorozmerného priestoru.

• <http://www.youtube.com/watch?v=9NrALgHFwTo&feature=related>

• <http://www.youtube.com/watch?v=3liCbRZPrZA>

Bernhard Schölkopf and Alexander Smola. *Learning with Kernels*. Adaptive Computation and Machine Learning. The MIT Press, Cambridge, Massachusetts, USA, 2002. ISBN 0-262-19475-9.

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^p$$

- Sa nazýva polynomiálny kernel stupňa p .
- Ak $p=2$, a vstupný vzor (vektor) má 7000 prvkov použitie kernela znamená vypočítať skalárny súčin so 7000 členmi a jeho následné umocnenie dvoma
- Ak by sme mali použiť explicitnú transformáciu do mnohorozmerného priestoru znamenalo by to výpočet približne 50,000,000 nových zaujímavých vlastností pre oba trénovacie vzory a následne počítať skalárny súčin uvedených vektorov
- Kernel trik teda umožňuje výrazným spôsobom znížovať výpočtovú náročnosť v porovnaní s explicitnou transformáciou

Multi-class SVM

oaasvm: One-Agains-All decomposition ,
 transforms the
 multi-class problem into a series of c
 binary subtasks

oaosvm: One-Against-One decomposition

$$g = c(c - 1)/2$$

transforms the multi-class problem into a
 series of g binary subtasks

Comparison in terms of accuracy on the
 digit dataset.

	error rate - no rejection -
MLP	0.80%
SVM - OAO	0.70%
SVM - OAA	0.63%

Comparison in terms of accuracy on the
 letter dataset.

	error rate - no rejection -
MLP	3.81%
SVM - OAO	3.22%
SVM - OAA	3.18%

Comparison of the two strategies in terms
 of complexity on the digit dataset.

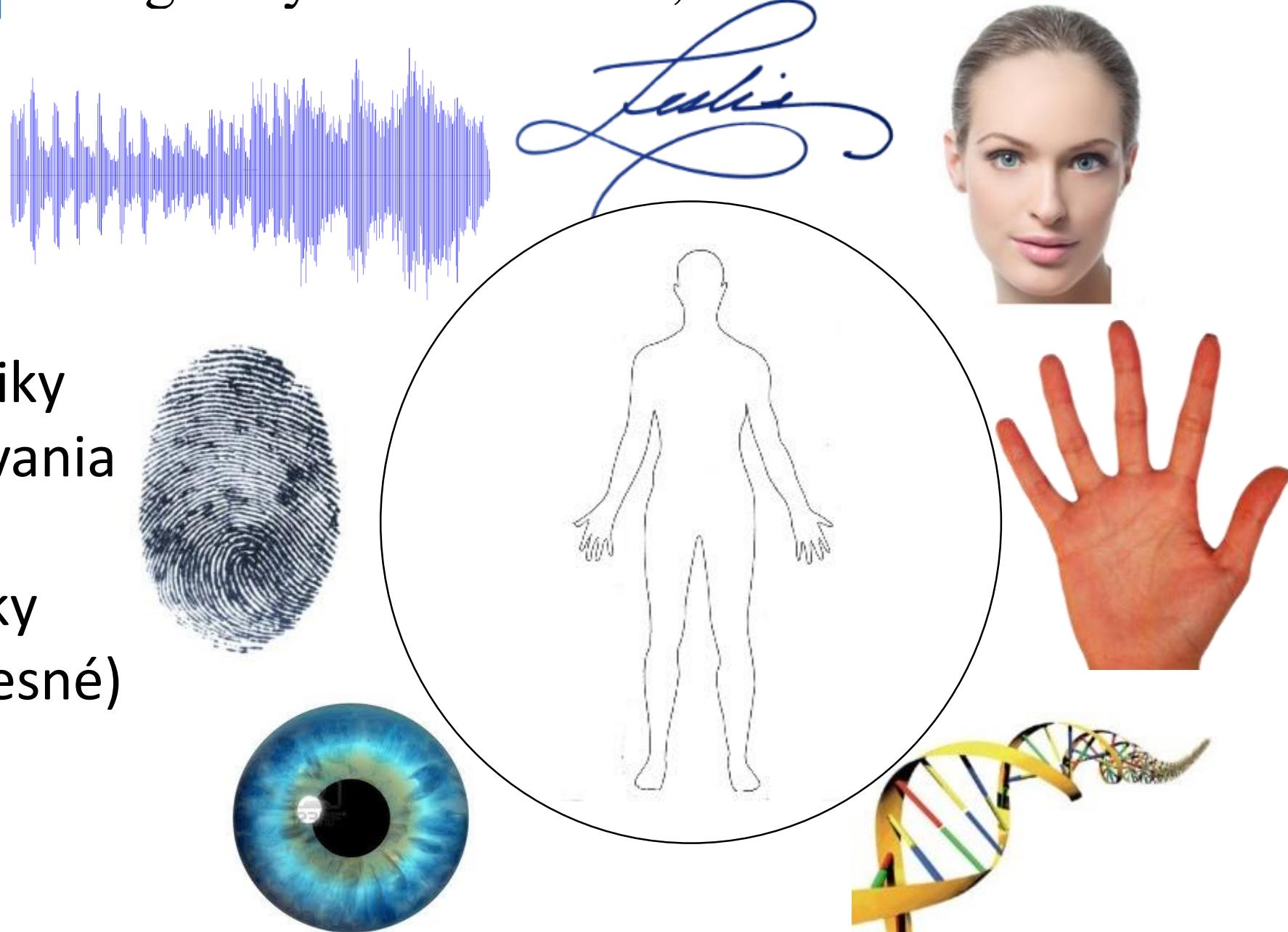
	OAO	OAA
number of SVMs	45	10
total training time	36 min.	32 h 17
number of SVs	5,753	8,514

Comparison of the two strategies in terms
 of complexity on the letter dataset.

	OAO	OAA
number of SVMs	325	26
total training time	4 min.	51 min.
number of SVs	9,152	11,109

Biometria

z gréckych: bio – život, metric – meranie



Behaviorálne biometriky

Biometriky správania

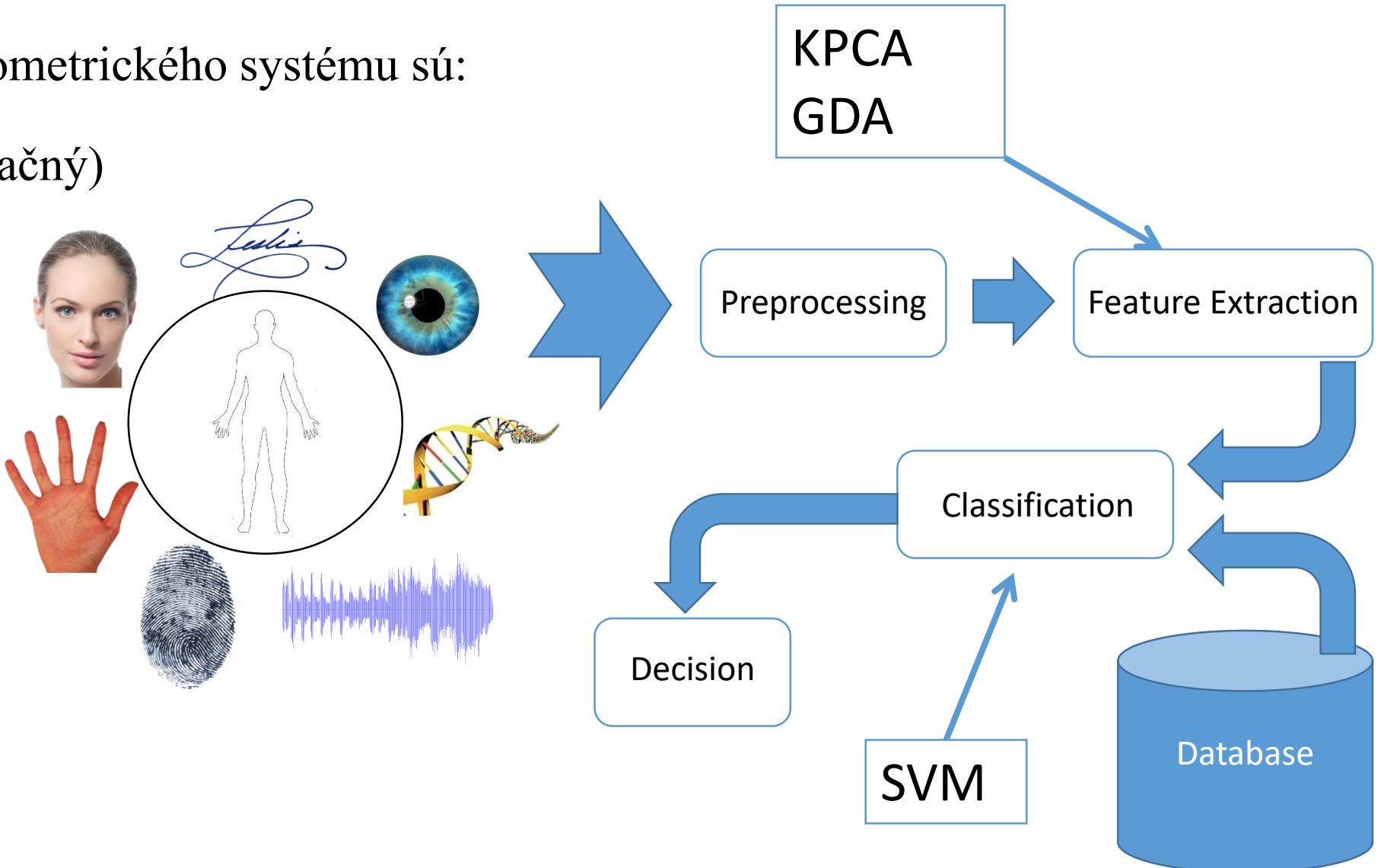
Fyziologické biometriky

Anatomické (telesné)
údaje

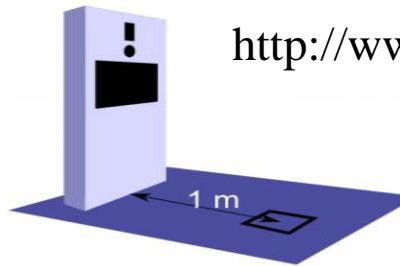
Biometrický rozpoznávací systém

Pracovné módy biometrického systému sú:

- Záznamový (registračný)
- Verifikačný
- identifikačný



Komerčné systémy - ľudská tvár



<http://www.xidtech.com/products.htm>

<http://www.samsung.com/us/experience/smart-tv/>

<http://www.lathem.com/products/automated-systems/fr700-face-recognition-system.aspx>



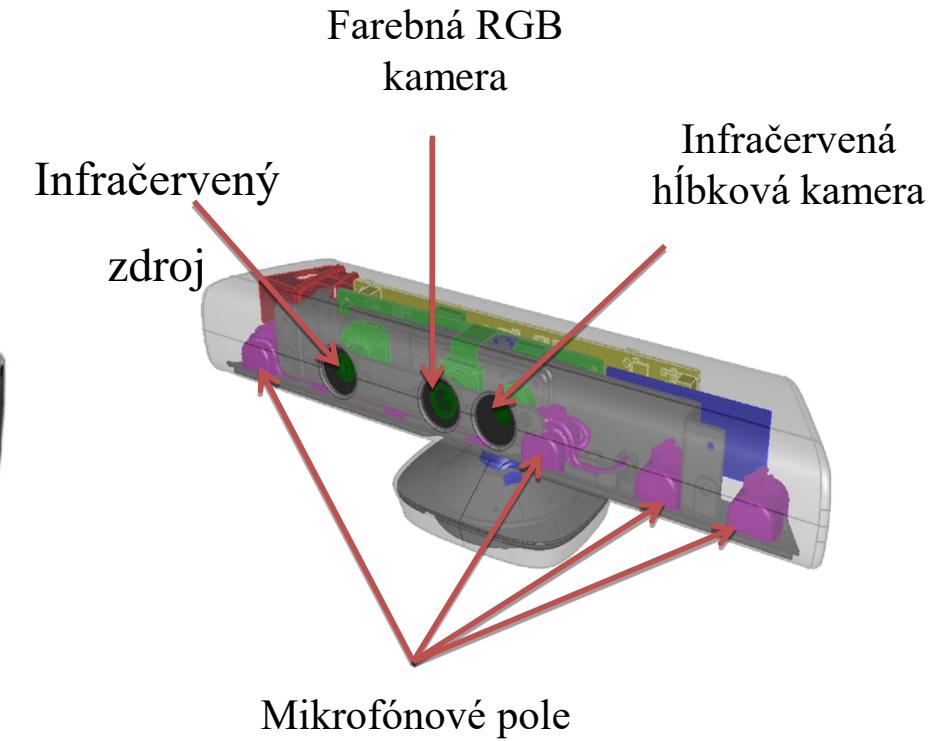
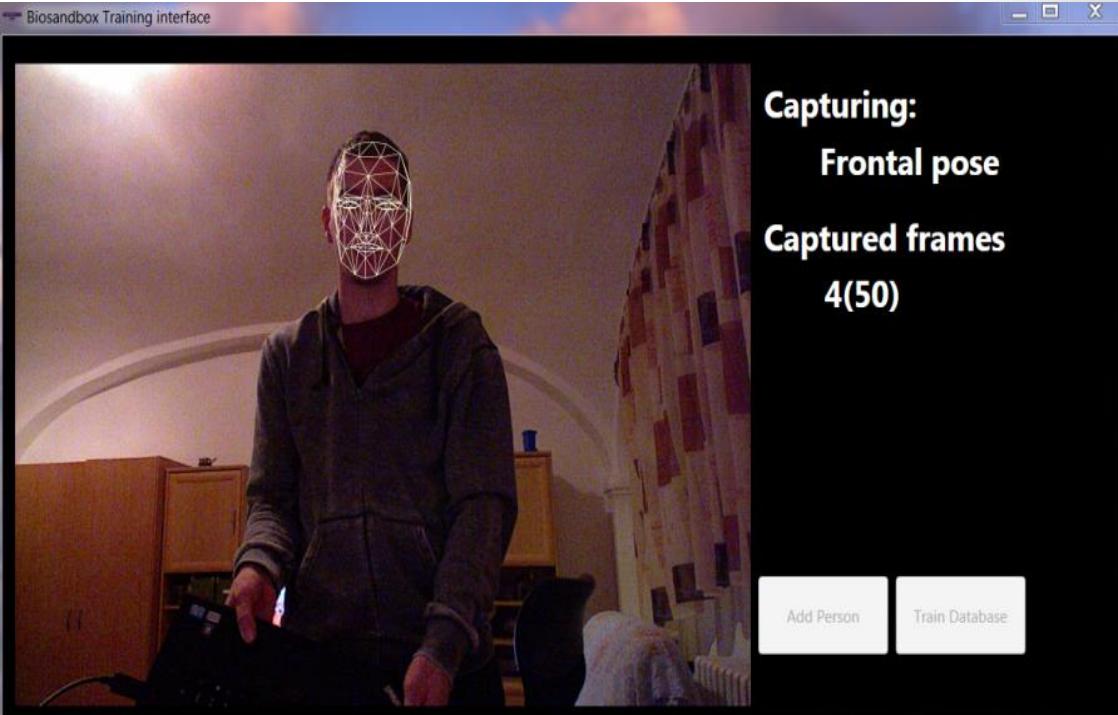
<https://www.youtube.com/watch?v=VA2jU7sXh-M>

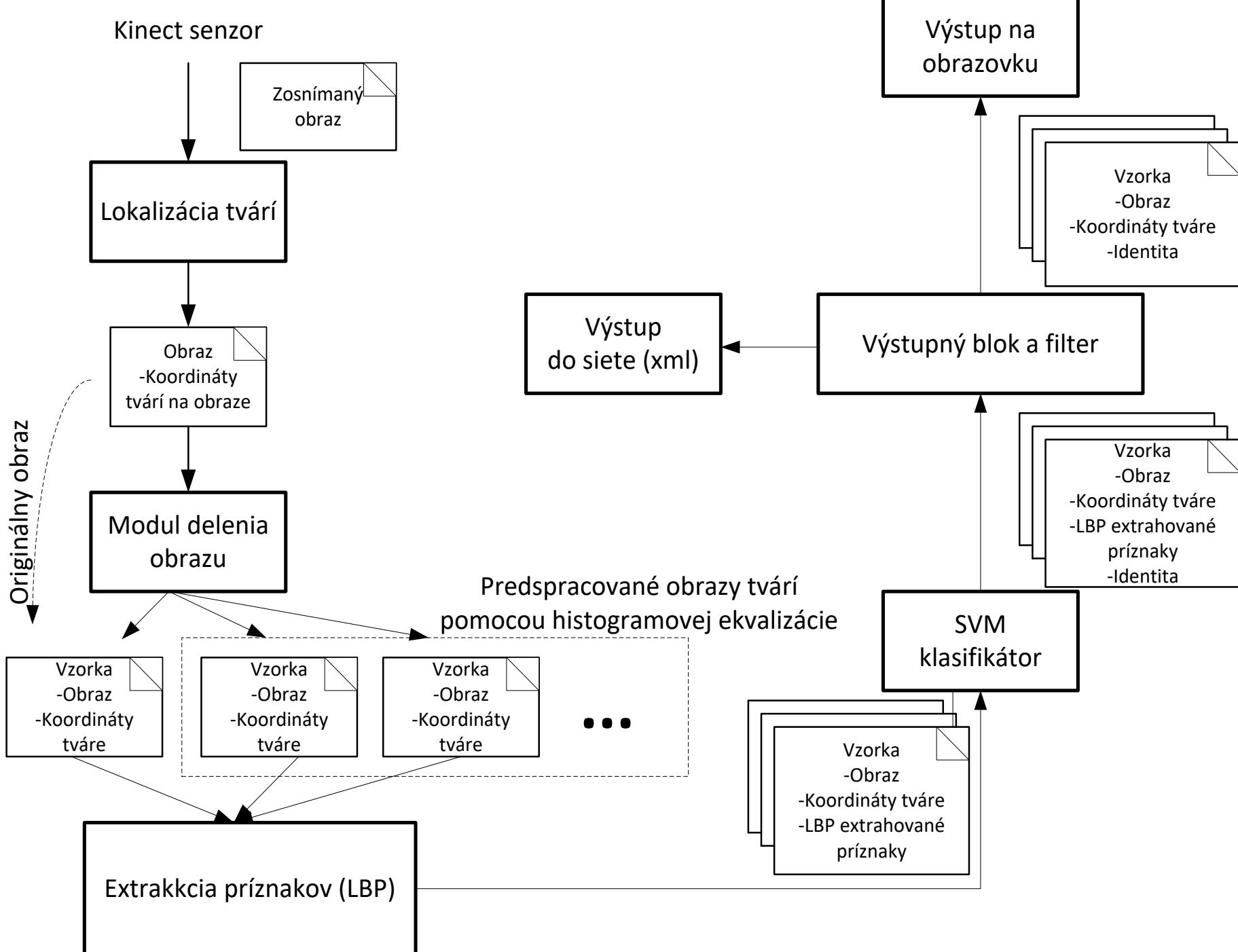
<https://www.youtube.com/watch?v=mdhvRNYX0PI>

https://www.youtube.com/watch?v=Ga0UCaoV8_o&list=PLa7iOxl-pL3leU2fmBAWRs993aldkshYp

<https://www.youtube.com/watch?v=sgEwBmGcE-s>

Biometrický rozpoznávací systém HBB NEXT





Matej_Feder



Matei Feder
Jozef_Ban



NEWS-2016

- <http://news.mit.edu/topic/machine-learning>
- <https://www.extremetech.com/tag/machine-learning>
- <http://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html>
- cloud.google.com/ml
- <https://dennikn.sk/558313/najziadanejsou-poziciou-na-svete-je-data-scientist>

NEWS-2015

- <http://www.theguardian.com/australia-news/2015/oct/21/facebook-photos-could-be-taken-for-use-in-national-biometric-database-officials>
- <http://www.planetbiometrics.com/article-details/i/3666/desc/dubai-airport-readies-for-lsquoiris-on-the-move/>
- <http://www.zive.sk/clanok/109061/presnejsie-navrhy-pri-pisani-vdaka-neuronovym-sietam-prinesie-ich-swiftkey>
- <http://www.infoworld.com/article/2901068/authentication/windows-10-biometric-security-front-and-center.html>
- <http://news.mit.edu/2015/predicting-change-alzheimer%E2%80%99s-brain-1006>
- <http://www.extremetech.com/extreme/211204-google-brings-deep-neural-networks-to-your-phone-with-translate>