

Klasifikátory

Zadanie č.2

Biometria

Obsah

- Zadanie č.2
 - Inštrukcie
 - Model riešenia a základné pojmy
 - Bodovanie

Inštrukcie

- Znenie zadania:
 - Softvérovo implementujte **2 klasifikátory**:
 1. Naivný Bayesovský klasifikátor
 2. SVM klasifikátor

Inštrukcie

- Znenie zadania:
 - Softvérovo implementujte **2 klasifikátory**:
 1. Naivný Bayesovský klasifikátor
 2. SVM klasifikátor
 - Otestujte obidva klasifikátory prostredníctvom „**cross-validácie**“

Inštrukcie

- Znenie zadania:
 - Softvérovo implementujte **2 klasifikátory**:
 1. Naivný Bayesovský klasifikátor
 2. SVM klasifikátor
 - Otestujte obidva klasifikátory prostredníctvom „**cross-validácie**“
 - Výstupom vášho programu bude **graf závislosti úspešnosti klasifikácie od počtu trénovacích vzoriek** (pre každý klasifikátor osobitne)

Inštrukcie

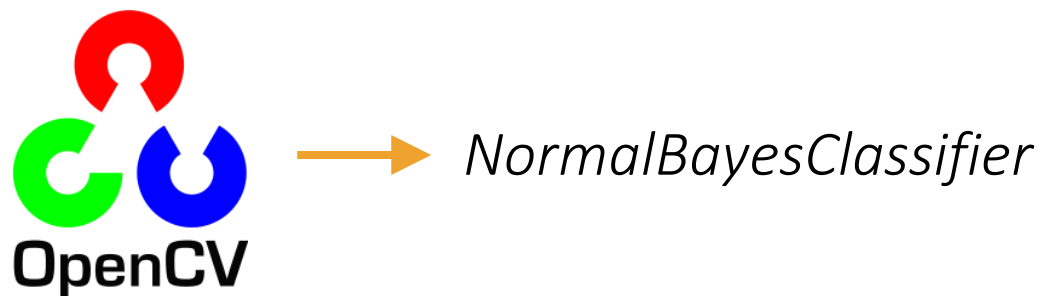
- Dôležitá poznámka:
 - Naivný Bayesovský klasifikátor **je nutné** implementovať „ručne“, teda bez akýchkoľvek externých knižníc

Inštrukcie

- Dôležitá poznámka:
 - Naivný Bayesovský klasifikátor **je nutné** implementovať „ručne“, teda bez akýchkoľvek externých knižníc
 - **Príklad:** nesmie sa použiť OpenCV modul, kde je Bayesovský klasifikátor už implementovaný v triede *NormalBayesClassifier*

Inštrukcie

- Dôležitá poznámka:
 - Naivný Bayesovský klasifikátor **je nutné** implementovať „ručne“, teda bez akýchkoľvek externých knižníc
 - **Príklad:** nesmie sa použiť OpenCV modul, kde je Bayesovský klasifikátor už implementovaný v triede *NormalBayesClassifier*



Inštrukcie

- Dôležitá poznámka:
 - Naivný Bayesovský klasifikátor **je nutné** implementovať „ručne“, teda bez akýchkoľvek externých knižníc
 - **Príklad:** nesmie sa použiť OpenCV modul, kde je Bayesovský klasifikátor už implementovaný v triede *NormalBayesClassifier*



NormalBayesClassifier



NaiveBayes



fitcnb()

Inštrukcie

- Dôležitá poznámka:
 - Naivný Bayesovský klasifikátor **je nutné** implementovať „ručne“, teda bez akýchkoľvek externých knižníc
 - **Príklad:** nesmie sa použiť OpenCV modul, kde je Bayesovský klasifikátor už implementovaný v triede *NormalBayesClassifier*



Zakázané

NormalBayesClassifier



MATLAB



NaiveBayes



fitcnb()

Model riešenia a základné pojmy

- Údaje na klasifikáciu
 - Úlohou oboch klasifikátorov je správne **klasifikovať písmená z abecedy**


Model riešenia a základné pojmy

- Údaje na klasifikáciu
 - Úlohou oboch klasifikátorov je správne **klasifikovať písmená z abecedy**
 - Je k dispozícii databáza 20 000 vzoriek, z ktorých každá popisuje konkrétne písmeno 17-prvkovým vektorom \mathbf{x}

Model riešenia a základné pojmy

- Údaje na klasifikáciu

- Úlohou oboch klasifikátorov je správne **klasifikovať písmená z abecedy**
- Je k dispozícii databáza 20 000 vzoriek, z ktorých každá popisuje konkrétne písmeno 17-prvkovým vektorom \mathbf{x}


- $\mathbf{x} = (\text{CLASS } X1 \ X2 \ X3 \ X4 \ X5 \ X6 \ X7 \ X8 \ X9 \ X10 \ X11 \ X12 \ X13 \ X14 \ X15 \ X16)$


trieda *atribúty*

Model riešenia a základné pojmy

- Údaje na klasifikáciu

- Úlohou oboch klasifikátorov je správne **klasifikovať písmená z abecedy**
- Je k dispozícii databáza 20 000 vzoriek, z ktorých každá popisuje konkrétne písmeno 17-prvkovým vektorom \mathbf{x}

- $\mathbf{x} = (\text{CLASS } X1 \ X2 \ X3 \ X4 \ X5 \ X6 \ X7 \ X8 \ X9 \ X10 \ X11 \ X12 \ X13 \ X14 \ X15 \ X16)$

trieda *atribúty*

- V databáze je zastúpená celá telegrafná abeceda (26 znakov)

Model riešenia a základné pojmy

- Údaje na klasifikáciu

- <https://archive.ics.uci.edu/ml/datasets/Letter+Recognition>



```
T, 2, 8, 3, 5, 1, 8, 13, 0, 6, 6, 10, 8, 0, 8, 0, 8  
I, 5, 12, 3, 7, 2, 10, 5, 5, 4, 13, 3, 9, 2, 8, 4, 10  
D, 4, 11, 6, 8, 6, 10, 6, 2, 6, 10, 3, 7, 3, 7, 3, 9  
N, 7, 11, 6, 6, 3, 5, 9, 4, 6, 4, 4, 10, 6, 10, 2, 8  
G, 2, 1, 3, 1, 1, 8, 6, 6, 6, 6, 5, 9, 1, 7, 5, 10  
S, 4, 11, 5, 8, 3, 8, 8, 6, 9, 5, 6, 6, 0, 8, 9, 7  
B, 4, 2, 5, 4, 4, 8, 7, 6, 6, 7, 6, 6, 2, 8, 7, 10  
A, 1, 1, 3, 2, 1, 8, 2, 2, 2, 8, 2, 8, 1, 6, 2, 7
```

Ukážka súboru „pismena.txt“

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Je klasifikátor založený na podmienenej („aposteriori“) pravdepodobnosti, ktorý aplikuje **Bayesovu teorému**.

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Je klasifikátor založený na podmienenej („aposteriori“) pravdepodobnosti, ktorý aplikuje **Bayesovu teorému**.
 - Majme vektor $\mathbf{x} = (x_1, x_2, \dots, x_n)$, ktorý treba klasifikovať do jednej z K tried.

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Je klasifikátor založený na podmienenej („aposteriori“) pravdepodobnosti, ktorý aplikuje **Bayesovu teorému**.
 - Majme vektor $\mathbf{x} = (x_1, x_2, \dots, x_n)$, ktorý treba klasifikovať do jednej z K tried.
 - Teda, potrebujeme vypočítať podmienenú pravdepodobnosť $p(C_K | x_1, x_2, \dots, x_n)$

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Je klasifikátor založený na podmienenej („aposteriori“) pravdepodobnosti, ktorý aplikuje **Bayesovu teorému**.
 - Majme vektor $\mathbf{x} = (x_1, x_2, \dots, x_n)$, ktorý treba klasifikovať do jednej z K tried.
 - Teda, potrebujeme vypočítať podmienenú pravdepodobnosť $p(C_K | x_1, x_2, \dots, x_n)$
 - Vypočítame ju pomocou Bayesovej teorémy:

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Je klasifikátor založený na podmienenej („aposteriori“) pravdepodobnosti, ktorý aplikuje **Bayesovu teorému**.
 - Majme vektor $\mathbf{x} = (x_1, x_2, \dots, x_n)$, ktorý treba klasifikovať do jednej z K tried.
 - Teda, potrebujeme vypočítať podmienenú pravdepodobnosť $p(C_K | x_1, x_2, \dots, x_n)$
 - Vypočítame ju pomocou Bayesovej teorémy:

$$p(C_K | \mathbf{x}) = \frac{p(C_K) p(\mathbf{x} | C_K)}{p(\mathbf{x})}$$

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Ako klasifikovať neznámu vzorku?

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Ako klasifikovať neznámu vzorku? Využije sa **MAP rozhodovacie pravidlo**.

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Ako klasifikovať neznámu vzorku? Využije sa **MAP rozhodovacie pravidlo**.
 - MAP = „maximum a posteriori“

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Ako klasifikovať neznámu vzorku? Využije sa **MAP rozhodovacie pravidlo**.
 - MAP = „maximum a posteriori“
 - Neznámej vzorke prideliť triedu \hat{C} nasledovne:

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor
 - Ako klasifikovať neznámu vzorku? Využije sa **MAP rozhodovacie pravidlo**.
 - MAP = „maximum a posteriori“
 - Neznámej vzorke pridelíme triedu \hat{C} nasledovne:

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k)$$

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k)$$

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k) \longrightarrow \hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k)$$

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k) \longrightarrow \hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} \boxed{A} \prod_{i=1}^n p(x_i | C_k)$$

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k) \longrightarrow \hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} \boxed{A} \prod_{i=1}^n \boxed{B}$$

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k) \longrightarrow \hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} \boxed{A} \prod_{i=1}^n \boxed{B}$$

A

Frekvencia triedy v trénovacej množine

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k) \longrightarrow \hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} \boxed{A} \prod_{i=1}^n \boxed{B}$$

A

Frekvencia triedy v trénovacej množine

B

Pravdepodobnostné rozdelenie atribútov vzorky v rámci triedy K

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

Frekvencia triedy v trénovacej množine

$$\boxed{A} = \frac{\text{počet výskytov vzoriek z triedy } k}{\text{veľkosť trénovacej databázy}}$$

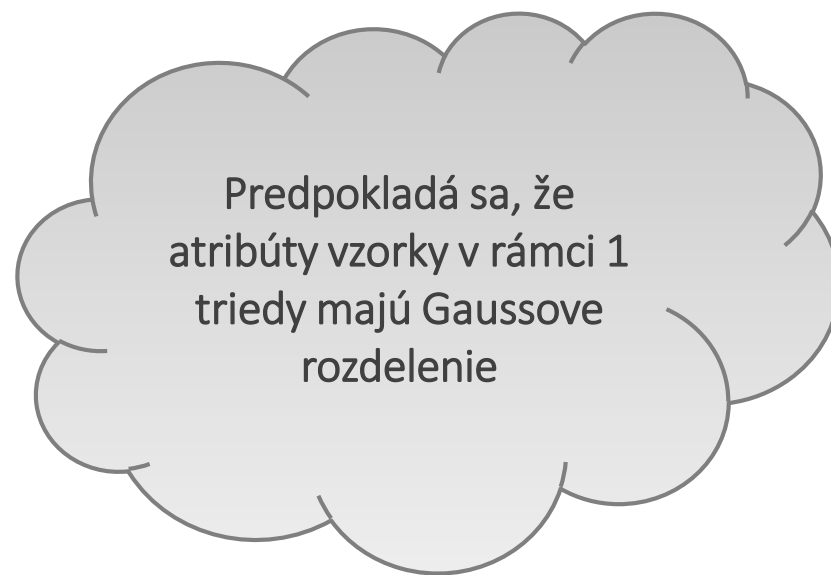
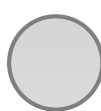
Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

Pravdepodobnostné rozdelenie atribútov x_i neznámej vzorky v rámci triedy K:

B

= ?



Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

Pravdepodobnostné rozdelenie atribútov x_i neznámej vzorky v rámci triedy K :

1. Rozdelíme vzorky v trénovacej databáze podľa tried

$$\boxed{B} = ?$$

Model riešenia a základné pojmy

○ Naivný Bayesovský klasifikátor

Pravdepodobnostné rozdelenie atribútov x_i neznámej vzorky v rámci triedy K:

B

= ?

1. Rozdelíme vzorky v trénovacej databáze podľa tried
2. Vypočítame priemer μ a rozptyl σ^2 pre každý atribút x_i v každej triede v trénovacej databáze

Model riešenia a základné pojmy

○ Naivný Bayesovský klasifikátor

Pravdepodobnostné rozdelenie atribútov x_i neznámej vzorky v rámci triedy K:

B

= ?

1. Rozdelíme vzorky v trénovacej databáze podľa tried
2. Vypočítame priemer μ a rozptyl σ^2 pre každý atribút x_i v každej triede v trénovacej databáze
3. Atribút x_i z neznámeho vektora \mathbf{x} , ktorý sa chystáme klasifikovať odovzdáme ako parameter do funkcie Gaussovho rozdelenia:

Model riešenia a základné pojmy

○ Naivný Bayesovský klasifikátor

Pravdepodobnostné rozdelenie atribútov x_i neznámej vzorky v rámci triedy K:

B

= ?

1. Rozdelíme vzorky v trénovacej databáze podľa tried
2. Vypočítame priemer μ a rozptyl σ^2 pre každý atribút x_i v každej triede v trénovacej databáze
3. Atribút x_i z neznámeho vektora \mathbf{x} , ktorý sa chystáme klasifikovať odovzdáme ako parameter do funkcie Gaussovho rozdelenia:

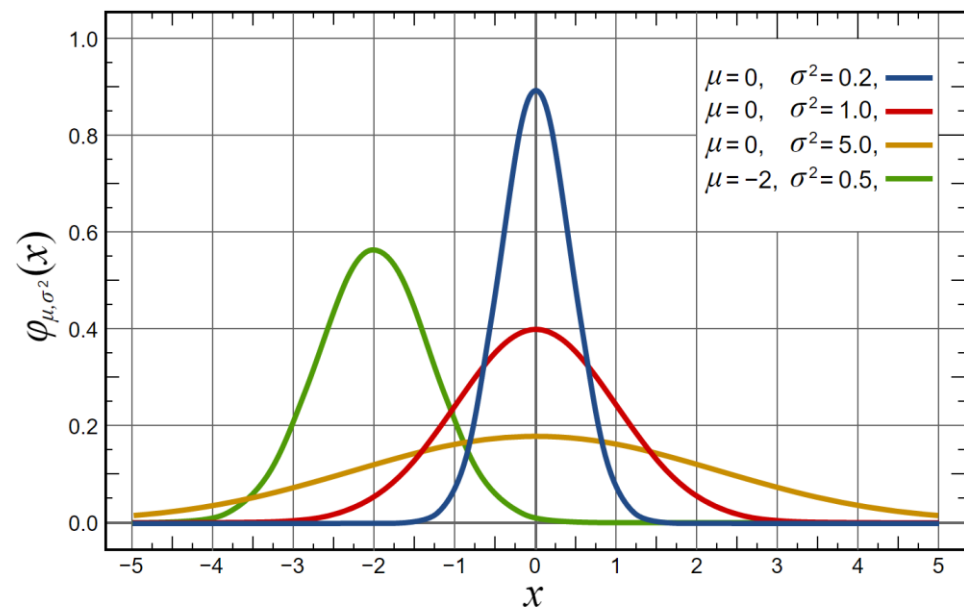
$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

Pravdepodobnostné rozdelenie atribútov x_i neznámej vzorky v rámci triedy K:

$$\text{B} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$



Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

Rekapitulácia

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} p(C_k) \prod_{i=1}^n p(x_i | C_k)$$

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

Rekapitulácia

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} \boxed{A} \prod_{i=1}^n p(x_i | C_k)$$

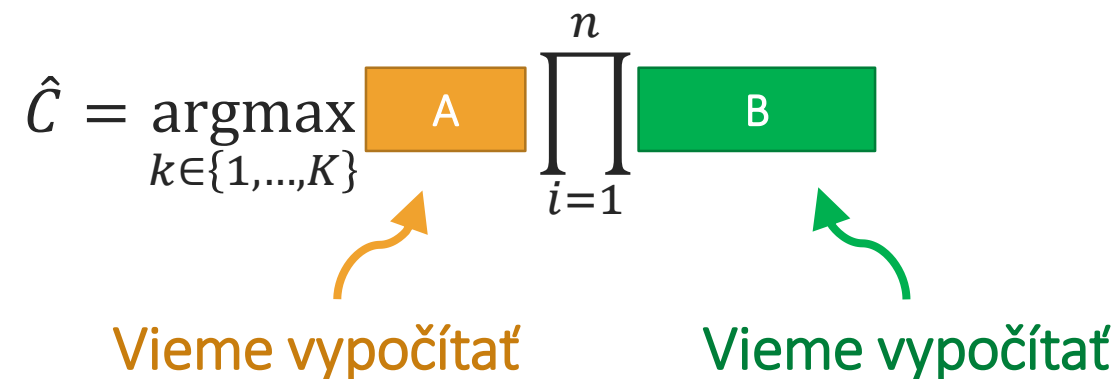
Vieme vypočítať

Model riešenia a základné pojmy

- Naivný Bayesovský klasifikátor

Rekapitulácia

$$\hat{C} = \operatorname{argmax}_{k \in \{1, \dots, K\}} \boxed{A} \prod_{i=1}^n \boxed{B}$$



Vieme vypočítať

Vieme vypočítať

Model riešenia a základné pojmy

- SVM klasifikátor

Pokračovanie nabudúce ...

Model riešenia a základné pojmy

- Výstupné grafy
 - Odovzdávajú sa 2 grafy:

Model riešenia a základné pojmy

- Výstupné grafy

- Odovzdávajú sa 2 grafy:

- 1. Úspešnosť naivného Bayesovského klasifikátora

Model riešenia a základné pojmy

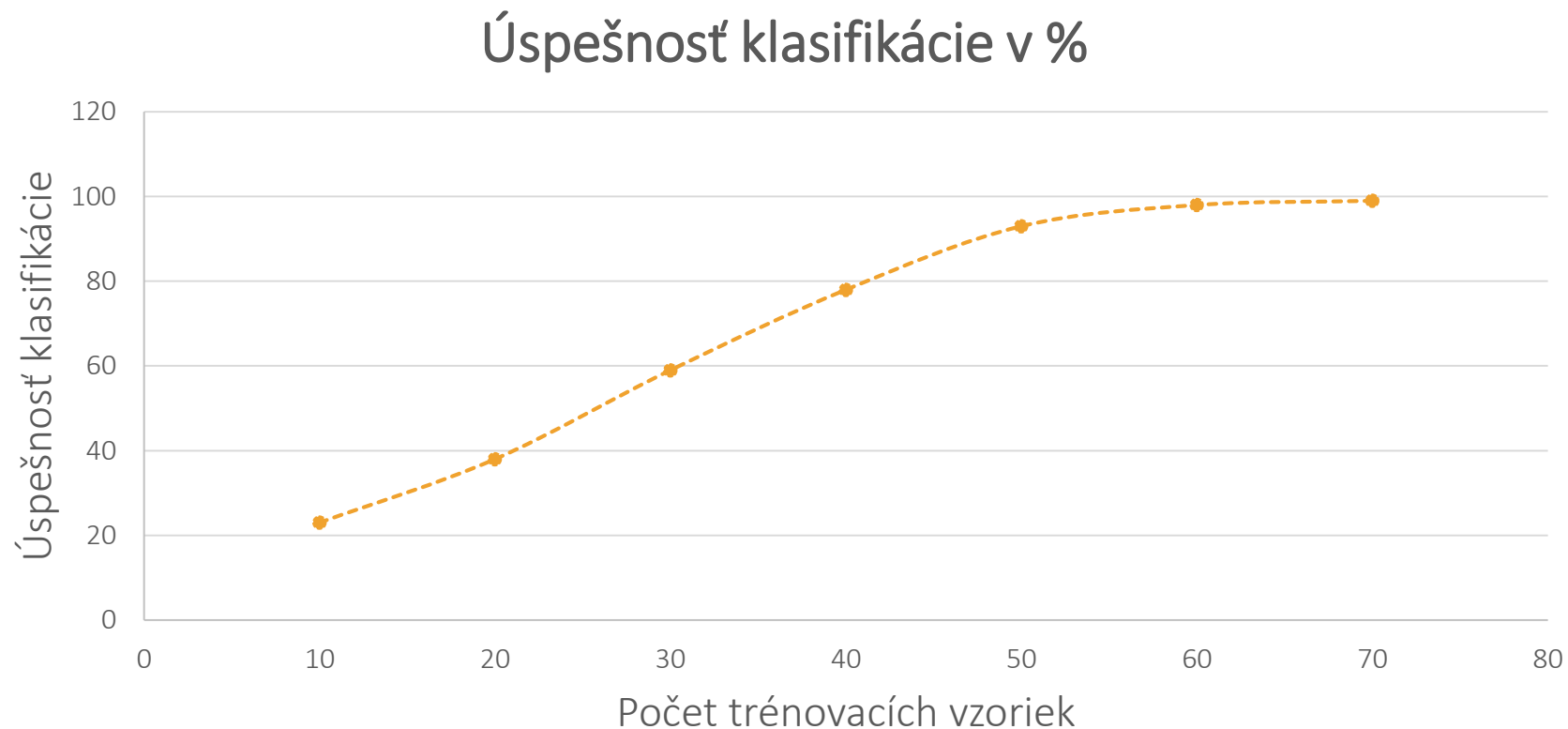
- Výstupné grafy

- Odovzdávajú sa 2 grafy:

1. Úspešnosť naivného Bayesovského klasifikátora
2. Úspešnosť SVM klasifikátora

Model riešenia a základné pojmy

- Vzor výstupného grafu



Bodovanie

Funkčnosť cross-validácie	2b
Vlastná implementácia naivného Bayesovského klasifikátora	4b
Grafy (2ks) s úspešnosťou klasifikácie	2b
Spolu	8b