

# Semestrální zkouška SRE, 16.1.2007

Jméno a příjmení: .....

Login: .....

Podpis: .....

**Příklad/Otázka 1** Jaký je důvod zavedení Gaussian-Mixture modelu, proč nám nestačí jedna Gaussovka ?

---

**Příklad/Otázka 2** Co je učící konstanta (learning rate) v algoritmu trénování neuronových sítí a jak by měla být optimálně nastavena (neopisujte prosím popis algoritmu NewBob...) ?

---

**Příklad/Otázka 3** Jaká je podstata supervised MAP-adaptace ?

---

**Příklad/Otázka 4** Jaký je hlavní rozdíl mezi Maximum likelihood (ML) a Maximum mutual information (MMI) trénováním ? Stačí, když se zaměříte na objektivní funkci.

---

**Příklad/Otázka 5** Základem při trénování HMM je výpočet “occupation counts”  $L_j(t)$ , které “měkce rozhazují” vektory na jednotlivé stavy. Uveďte, proč při jejich výpočtu nevystačíme s prostými vysílacími likelihoody  $b_j(\mathbf{o}(t))$ , ale proč potřebujeme vyhodnocení pomocí

$$L_j(t) = \frac{p(\mathbf{O}, x(t) = j|M)}{\sum_j p(\mathbf{O}, x(t) = j|M)}.$$



**Příklad/Otázka 10** S fonémovým rozpoznávačem založeným na rozděleném časovém kontextu (split temporal context - STC) dosahujeme lepších výsledků než s celou časovou trajektorií, i když je struktura rozpoznávače složitější (3 neuronové sítě místo jedné). Uveďte stručně, jaký je pro zavedení STC důvod.

---

**Příklad/Otázka 11** Hodnoty jednoho prvku feature-vektoru na trénovacím setu byly:

[12 13 10 0.1 0.4 5 9 23 0.9 4.5 5.2]

Určete, jakou hodnotu bude mít tento prvek pro 4.5 po **rank-normalizaci**.

---

**Příklad/Otázka 12** Popište stručně eigen-channel adaptaci v rozpoznávání mluvčího.

---

**Příklad/Otázka 13** Který způsob identifikace jazyka (LID) by byl podle Vás vhodnější pro rozpoznávání pražského a brněnského dialektu a proč? Nemusíte se omezit pouze na přístupy prezentované na přednášce!

---

**Příklad/Otázka 14** Definovali jsme několik požadavků na “dobré” koeficienty pro klasifikátory založené na GMM. Shifted delta cepstra (SDC) používané v identifikaci jazyka (LID) jeden z těchto požadavků drsným způsobem porušují. Vysvětlete, o který požadavek jde, a proč je porušen.

**Příklad/Otázka 15,16** Mějme množinu A jednoduchých vět:

1. ko ko dák
2. ko ko ko dák
3. kvo kvo ko dák
4. ko ko
5. kvo ko ko
6. ko ko dák
7. kvo ko ko dák

a množinu B jednoduchých vět:

1. kvo kvo
2. ko ko

Určete perplexitu nejjednoduššího modelu nultého řádu (unigramový model — pravděpodobnosti slov) vypočteného z trénovací množiny A na testovací množině B.

---

**Příklad/Otázka 17,18** Použijte stejná data jako v minulém příkladu. Označme:

- $M_0$  nejjednodušší model nultého řádu (unigramový model — pravděpodobnosti slov) vypočtený pomocí MLE z trénovací množiny A.
  - $M_1$  nejjednodušší model prvního řádu (bigramový model — pravděpodobnosti dvojic slov + Začátek + Konec) vypočtený pomocí MLE z trénovací množiny A.
1. Určete a porovnejte pravděpodobnosti slepičích “vět” *ko ko ko* a *ko ko dák* při použití modelu  $M_0$ .
  2. Určete a porovnejte pravděpodobnosti slepičích “vět” *ko ko ko* a *ko ko dák* při použití modelu  $M_1$ .
  3. Jaká je pravděpodobnost “věty” *kvo dák* při použití modelu  $M_0$  a jaká při použití modelu  $M_1$ ?

**Příklad/Otázka 19** Určete pravděpodobnost překladu *Spisovatel psal o tobě* v překladovém modelu při překladu z angličtiny do češtiny (použijte odhad maximální věrohodnosti), pokud je v trénovacích paralelních textech:

this	ten	2 ×
the	ten	5 ×
the	tento	1 ×
the		16 ×
writer	spisovatel	8 ×
has	má	5 ×
has		5 ×
written	napsaný	2 ×
written	psal	3 ×
write	píše	6 ×
wrote	psal	7 ×
written	píše	1 ×
about	o	20 ×
about	na	5 ×
at	o	3 ×
a		30 ×
a	nějaký	3 ×
you	tobě	2 ×
you	ty	10 ×
yourself	tobě	1 ×
yourself	ty	2 ×

a slova (vč. prázdného překladu) uvedená v tabulce se jinde nevyskytují.

**Příklad/Otázka 20** Určete pravděpodobnost věty *Novák kupuje pole u potoka* s pravděpodobnostní bezkontextovou gramatikou:

$S \rightarrow NP VP$	1,0
$NP \rightarrow NProp$	0,2
$NP \rightarrow N$	0,7
$NP \rightarrow N PP$	0,1
$VP \rightarrow V NP PP$	0,4
$VP \rightarrow V NP$	0,6
$PP \rightarrow PREP N$	1,0
$NProp \rightarrow Novák$	1,0
$N \rightarrow pole$	0,5
$N \rightarrow potoka$	0,5
$V \rightarrow kupuje$	1,0
$PREP \rightarrow u$	1,0

