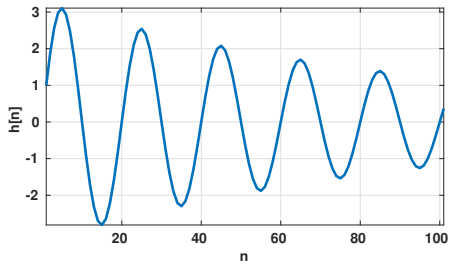
Semestrální zkouška ZRE, řádný termín, 18.5.2021, skupina Nicedaŋ

Login: ..... Příjmení a jméno: ..... Podpis: .....
(prosím čitelně!)

1. Řečový signál $x[n]$ přichází online a je potřeba ho ihned zpracovat. Popište, jak v tomto scénáři naimplementujete jeho ustřednění.

---

2. Impulsní odezva filtru IIR druhého řádu má tvar jen velmi pomalu se zeslabujícího harmonického signálu o periodě $N = 20$ vzorků. Nakreslete polohy pólů přenosové funkce tohoto filtru v rovině z.



---

3. Nakreslete modulové spektrum znělého úseku řeči - hlásky "i". Frekvence základního tónu je F_0 =100 Hz, frekvence dvou hlavních formantů jsou F_1 =240 Hz, F_2 =2400 Hz. Vzorkovací frekvence je F_s =8000 Hz, spektrum kreslete jen od nuly do poloviny F_s.

---

4. Rámec řeči má 160 vzorků a pouze jeden z nich má hodnotu 10, ostatní jsou nula. Určete, jak budou vypadat DFT-cepstrální koeficienty tohoto rámce. Nultý cepstrální koeficient $c[0]$ neuvažujte.
Pomůcka: $c[n] = \mathcal{F}^{-1}\left\{\ln|\mathcal{F}\{x[n]\}|^2\right\}$, přímá Fourierova transformace je definována pomocí DFT: $X[k] = \sum_n x[n]e^{-j\frac{2\pi}{N}nk}$ a zpětná pomocí IDFT: $x[n] = \sum_k X[k]e^{j\frac{2\pi}{N}nk}$.

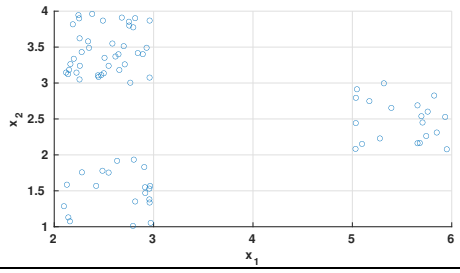
---

5. Je dáno 8 vzorků signálu $x[n]$: [1 0 -1 0 1 0 -1 0].
Určete koeficienty a_1 a a_2 prediktoru druhého řádu. Správnost výpočtu zkontrolujte pomocí intuice "jak mají vypadat koeficienty a_1 a a_2 filtru $1 - A(z) = -a_1 z^{-1} - a_2 z^{-2}$, aby tento filtr co nejlépe predikoval signál $x[n]$?

6. Používá se v kódování řeči (např. produkční kodeky typu CELP) chybový signál ("reziduál") lineární predikce ? Pokud ano, jak ?

---

7. Na obrázku je 80 trénovacích vektorů $\mathbf{x}[n]$. Abyste je nemuseli počítat, v každém clusteru je jich 20, vlevo nahoře 40. Nakreslete pozice kódových vektorů $\mathbf{y}_i$ kódové knihy vektorového kvantování (VQ) s $L = 4$ kódovými vektory. **Přibližně** určete totální vzdálenost této kódové knihy při kódování trénovacích dat: $D_{VQ} = \frac{1}{N} \sum_{n=1}^{N} d\left(\mathbf{x}[n], \mathbf{y}_i[n]\right)$, kde $\mathbf{y}_i[n]$ je nejbliší kódový vektor k trénovacímu vektoru $\mathbf{x}[n]$. Jako vzdálenost $d(\cdot, \cdot)$ použijte běžnou Euklidovu vzdálenost.



---

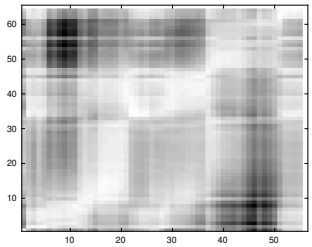
8. Proč je velikost normalizačního faktoru v klasické metodě dynamického borcení času (DTW) $N = R + T$ (kde R je počet vektorů referenční promluvy a T je počet vektorů testovací promluvy) konstantní pro všechny možné cesty ?

---

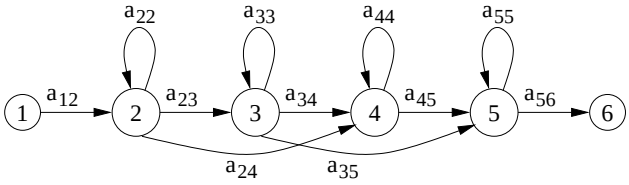
9. Vysvětlete, co je to u DTW "back-tracing".

---

10. Na obrázku je matice lokálních vzdáleností vektorů ("každý s každým") pro výpočet DTW. Menší vzdálenosti jsou zobrazeny jako světlejší. Nakreslete do obrázku přibližný průběh optimální srovnávací cesty.



11. Srytý Markovův model (HMM) na obrázku má reprezentovat promluvu o délce $T = 3$ feature vektory. Napište všechny možné stavové sekvence X. Uvědomte si, že v každé sekvenci musí být stav č. 1 na začátku a stav č. 6 na konci. Tyto dva stavy nereprezentují žádný vektor.



---

12. Je definován levo-pravý HMM se čtyřmi stavy, z toho 2 vysílací, přechodové log. pravděpodobnosti jsou:
$\log a_{12} = 0, \quad \log a_{22} = -0.51, \quad \log a_{23} = -0.92, \quad \log a_{33} = -0.36, \quad \log a_{34} = -1.2.$
Tabulka logaritmů hodnot funkcí hustoty vysílacích likelihoodů je:

t	...	46	47	48	...
$\log b_2(\mathbf{x}_t)$	...	-1	-2	-3	...
$\log b_3(\mathbf{x}_t)$	...	-4	-5	-6	...

Provádíme Viterbiho algoritmus pomocí "token passing". Hodnota tokenu ve stavu 2 v čase 46 je $\Psi_2(46) = -21$. Určete hodnotu tokenu ve stavu 2 v čase 48.

$\Psi_2(48) =$ .....

---

13. Uvažujte skrytý Markovův model, kde jsou funkce hustoty vysílacích likelihoodů modelovány pomocí Gaussovek. Uvažujte pouze jednu trénovací promluvu. Popište velmi stručně, jak se odhadnou parametry Gaussovek. Zaměřte se na způsob, jak jsou jednotlivé trénovací feature-vektory rozdělovány stavům.

---

14. Při rozpoznávání řeči pomocí váhovaných konečných stavových převodníků (wFST) je výsledná rozpoznávací síť dána jako $HCLG = H \circ C \circ L \circ G$. Napište stručně význam symbolů H, C, L, G.

---

15. Jedním z možných výstupů rozpoznávače řeči s velkým slovníkem může být tzv. word lattice (slovní mřížka). Popište, co v ní najdeme a/nebo malou lattici nakreslete.

16. V rozpoznávání řeči pomocí neuronových sítí se v poslední vrstvě používá speciální nelinearita, která zajišťuje, že výstupy budou dobře reprezentovat **pravděpodobnosti** jednotlivých jednotek (např. fonémů). Napište, jak se nelinearita jmenuje a napište její rovnici. Pokud ji neznáte, vymyslete ji !

---

17. Máte k disposici modul, který ze zvukového souboru vypočítá vektorovou reprezentaci s nízkou fixní dimenzionalitou (embedding), např. extraktor i-vektorů nebo x-vektorů. Jak pro dva embeddingy dostanete skóre udávající podobnost mluvčích ?

---

18. Co je při vyhodnocování systémů pro verifikaci řečníka (a mimochodem jakýchkoliv jiných detektorů) **equal error rate** (EER) ?

---

19. Systém pro dvojici nahrávek udává skóre, že je v nich ten samý mluvčí. Uveďte, jak nastavíte rozhodovací práh pro tvrdé rozhodnutí "stejný mluvčí" vs. "různý mluvčí" pro (1) aplikaci pro přístup k bankovnímu účtu pomocí hlasu (2) aplikaci pro vyhledání hlasu teroristy ve 100000 videích na Youtube.

---

20. Co v systémech pro syntézu řeči z textu (TTS) znamená "generování prozodie" ?