



۱. برای سیستم‌های زیر، کارکرد سیستم و داده‌های ورودی و خروجی را توصیف کنید:

- سامانه مکالمه گفتاری (Spoken dialogue system)
- تشخیص فعالیت گفتار (Voice activity detection)

۲. درست یا غلط بودن عبارت زیر را با یا مشخص کنید.

۱. ---- قضیه پارسوال بیان می‌کند که انرژی سیگنال در حوزه زمان و فرکانس با هم برابرند.
۲. ---- در فرایند تولید واکه‌ها (Vowel) تارهای صوتی نوسان می‌کنند.
۳. ---- افزایش تعداد لایه‌های شبکه عصبی همواره منجر به افزایش قابلیت تعمیم می‌شود.
۴. ---- چندی‌سازی (Quantization) به معنی گسسته کردن دامنه یک سیگنال آنالوگ است.

۳. تفاوت روش‌های تخمین بیشینه شباهت (ML) و بیشینه احتمال پسین (MAP) چیست؟

۴. تخمین بیشینه شباهت (MLE) را برای پارامترهای a, b با فرض داشتن تعداد N نمونه از داده‌هایی با تابع توزیع $p(x) = e^{-\left(\frac{x-a}{b}\right)^2}$ حساب کنید.

۵. تبدیل فوریه گسسته (DFT) سیگنال $x[n] = \delta[n]$ را بدست آورید. از مقدار حاصل، تبدیل معکوس فوریه بگیرد. آیا سیگنال اولیه حاصل می‌شود؟

۶. فرض کنید یک فایل صوتی حاوی عبارت "ایران زیبا" را از صدای خود ضبط کرده‌اید. نحوه محاسبه فرکانس گام (Pitch) صدای خود را به صورت گام به گام با یکی از روش‌های آن بیان کنید.

۷. می‌خواهیم برای تشخیص زبان (Language Identification) از روی صدا از یک شبکه عصبی استفاده

کنیم. برای این کار داده‌های مربوط به ۱۰ زبان مختلف را در اختیار داریم.

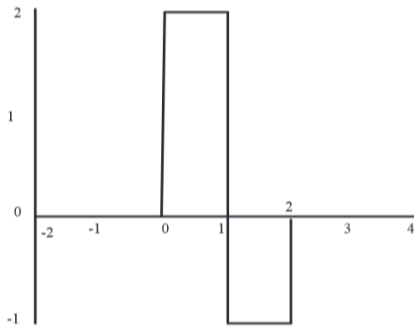
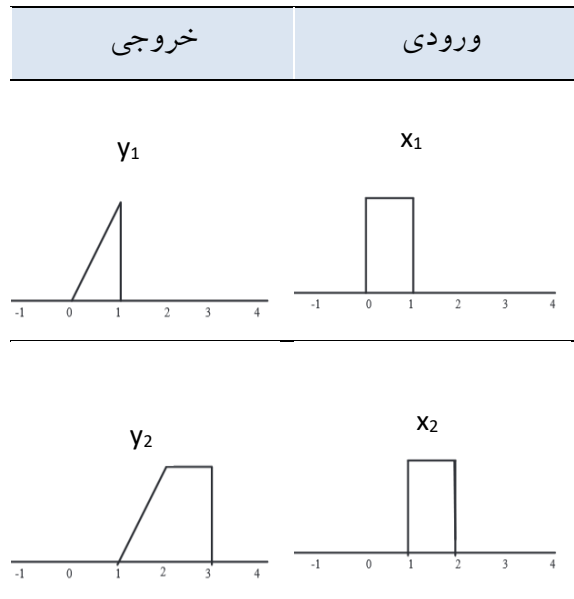
الف) یک نوع شبکه (MLP, LSTM, ...) برای این کار انتخاب کنید. علت انتخاب خود را بیان کنید.

ب) ساختار شبکه را به صورت کامل (شامل تعداد نرون‌های ورودی و خروجی) رسم کنید.

ج) چگونه می‌توان از شبکه CNN برای بهبود کارایی روش پیشنهادی خود استفاده کنید؟



۸. ورودی‌ها و خروجی معادل یک سیستم خطی به صورت زیر است



الف) آیا این سیستم نامتغیر با زمان (Time-Invariant)

است؟ چرا؟

ب) پاسخ سیستم را برای ورودی روبرو رسم کنید.