

**مسئله:** در این مسئله هدف استفاده از برنامه نویسی نزدیکترین همسایه در دسته بندی بیماری غده تیروئید است. داده مورد استفاده دارای 100 مثال است که هر کدام شامل 7 مشخصه و یک خروجی میباشد. خروجی میتواند یکی از مقادیر 0 معادل عدم تشخیص بیماری 1، 2 یا 3 تشخیص بیماری با یکی از سه کلاس مختلف باشد. مشخصه ها نیز به ترتیب عبارتند از:

- Age
- ON THYROXINE
- TSH MEASURED
- TSH
- T3
- TT4 MEASURED
- TT4

که مشخصه های 2، 3 و 6 باینری و باقی مشخصه ها حقیقی هستند.

داده مورد نظر را به 5 بخش مساوی تقسیم کنید و با استفاده از Validation-Cross اثر تغییر k از مقدار 1 تا 10 را بر روی متوسط خطای دسته بندی در الگوریتم KNN محاسبه نمایید.

این کار را در دو حالت انجام میدهیم:

الف) حالتی که داده غیر نرمالیزه وارد الگوریتم میگردد

ب) حالتی که داده به گونه ای نرمالیزه میشود که میانگین هر یک از مشخصه ها صفر و واریانس آنها واحد میگردد.