

سوال ۷) با توجه به اینکه الگوریتم به نقاط شروع وابسته است برای اینکه هر سری با نقاط شروع متفاوتی الگوریتم اجرا شود مقدار پارامتر `random state` این الگوریتم را با مقادیر متفاوت ۱ تا ۱۰ برای هر تعداد خوشه در نظر گرفته و در هر مرحله هر تعداد که خطای کمتری دارد را میان این ۱۰ مقدار انتخاب کردیم. در نهایت برای هر تعداد خوشه یک مقدار خطا و همچنین یک مقدار سیلوئنتی بدست آمد که البو و نمودار سیلوئنت رسم شد. با آنالیز نمودار البو متوجه میشویم در نقطه ۴ و ۶ یک شکستگی با شیب کندی وجود دارد که در نقطه ۴ این میزان شیب بیشتر است. مقدار سیلوئنت نیز در نقطه ۶ بیشتر از مابقی است. در ادامه نیز تعداد ۴ تا خوشه را در نظر گرفتیم.

سوال ۱۱) الگوریتم با هر دو مقدار خوشه ۴ و ۶ روی داده های کاهش بعد یافته اجرا کرده و در نهایت مقادیر خروجی با تعداد ۴ تا خوشه انتخاب شد.