

سؤال ۱:

$$w_i = w_i + \alpha * 1/m * \sum_{i=1}^m (y_i - hw) * x_i \quad ۱.$$

۲. بدون L2-norm بر خلاف پایینی نتیجه بهتری روی داده‌های train داد! برای جلوگیری از overfitting از regularization استفاده میشود، روش‌های مختلفی دارد از جمله L2.

۳. در logistic بدون L2 خطا و تابع هزینه کمتر بود و با L2 بیشتر بود، زیرا L2 می‌آید و overfitting را کاهش می‌دهد در نتیجه روی خود داده‌های train ممکن است حتی نتیجه بدتری هم دهد.

$$hw = \text{sigmoid}\left(\sum_{i=1}^n w_i * x_i\right) = 1 / (1 + e^{-(\sum_{i=1}^n x_i * w_i)})$$

اگر

$$w_i = w_i + \alpha * \sum_{i=1}^m (y_i - hw) * x_i$$

باشد، آنگاه:

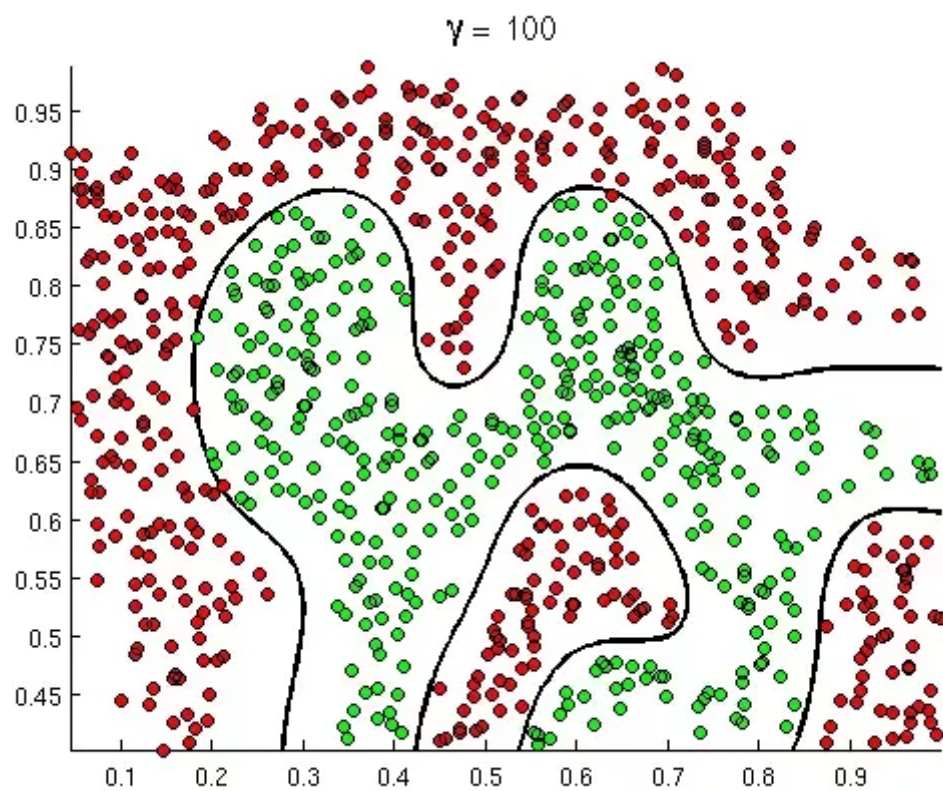
$$F = (w_0 + w_1 + x) * (-1/w_2) \text{ و } w \text{ همان } \theta \text{ است.}$$

Logistic برای classification و داده‌های گسسته است ولی linear برای regression و داده‌های پیوسته.

۴. <https://www.edvancer.in/logistic-regression-vs-decision-trees-vs-svm-part1>

<https://vitalflux.com/machine-learning-use-logistic-regression-vs-svm>

هر دو که اولاً برای classification هستند. روش logistic خط جداکننده، خط صاف است و خطی است (خوب نیست ممکن است داده‌ها غیر خطی باشند، مثلاً جداکننده دایره باشد که داخل یک label و خارج یک label دارد) بنابراین برای داده‌های خطی جداپذیر خوب است ولی svm با kernel کردن بعد داده‌ها را زیاد میکند و در آنجا خطی جدا میکند و با برگشتن از آنجا به بعد اصلی سؤال خط جداکننده دیگر خطی نیست. مثلاً شکل زیر یک نمونه از جداکننده‌ی svm است.



از طرفی چون logistic احتمالی است اگر داده‌ها خطی جداپذیر باشند شاید بهتر باشد که از آن استفاده کنیم ولی در مجموع svm بهتر است. همچنین در مساله ما svm سریعتر بود.