

یک نکته برای حل مشکل محاسبات ضرب اعداد کوچک در مسئله نایو بیز

ممکن است به این مشکل برخورد داشته باشید که زمانی که می‌خواهید دو احتمال $P(Y=+1|x_{\text{new}})$ و $P(Y=-1|x_{\text{new}})$ را محاسبه و با هم مقایسه کنید احتمالات انقدر کوچک باشند که برنامه شما امکان محاسبه با آن دقت را نداشته باشد و مقدار صفر را به عنوان ضرب احتمالات برگرداند. همان طور که می‌دانید برای محاسبه این احتمال در Naive Bayes classifier از رابطه زیر استفاده می‌کردیم .

$$\arg \max_y p(y|x) = \arg \max_y \frac{p(x|y)p(y)}{p(x)} = \arg \max_y p(x|y)p(y).$$

همچنین :

$$\begin{aligned} p(x_1, \dots, x_{50000}|y) \\ &= p(x_1|y)p(x_2|y)p(x_3|y) \cdots p(x_{50000}|y) \\ &= \prod_{i=1}^n p(x_i|y) \end{aligned}$$

پس:

$$\arg \max_y p(y|x) = \arg \max_y \prod_{i=1}^n (p(x_i|y)) * p(y)$$

حال قبل از مقایسه از دو طرف لگاریتم می‌گیریم در مبنای ۲ یا مبنای دلخواه بزرگتر از ۱ چون تابع لگاریتم اکیدا صعودی است جهت نامساوی عوض نخواهد شد . و در ضمن ضرب داخل لگاریتم به جمع تبدیل خواهد شد . بدین صورت اعداد کوچک نزدیک به صفر تبدیل به عدد منفی با اندازه بزرگتر شده و دیگر نیاز به ضرب نیز نداریم و جمع لگاریتم احتمالات را باهم مقایسه می‌کنیم .

$$\log(P(A) * P(B) * P(C)) = \log(P(A)) + \log(P(B)) + \log(P(C))$$

موفق باشید