Машинное обучение: Лекция 2

Гаврилов Даниил

2017-2018

olavursky@gmail.com
Санкт-Петербургский Государственный Университет
Факультет Прикладной Математики и Процессов Управления

Параметризация модели

Параметризация модели

Введем параметризацию некоторой модели, которая предсказывала бы нам цену квартиры

$$y = Xw + b, w \in \mathbb{R}^k, b \in \mathbb{R}$$

Очевидно, что такая это параметризация может быть описана как

$$p(y|X, w, b) = \mathcal{N}(y|Xw + b, \sigma^2)$$

Пеереобозначим w и b как $\theta \in \mathbb{R}^{k+1}$
X – матрица $N \times k$, где k – колличество "фичей"данных, а N – кол-во наблюдений,
y – вектор $\in \mathbb{R}^N$

Maximum likelihood

Самым простым способом обучить такую модель будет найти максимум правдоподобия

$$\theta_{ML} = argmax_{\theta}\left[p(y|X,\theta)\right] = argmax_{\theta}\left[ln\left(p(y|X,\theta)\right)\right]$$

Maximum likelihood

Самым простым способом обучить такую модель будет найти максимум правдоподобия

$$\theta_{ML} = argmax_{\theta}\big[p(y|X,\theta)\big] = argmax_{\theta}\Big[ln\big(p(y|X,\theta)\big)\Big]$$

Как и говорили в прошлый раз, если $p(y|x,\theta) = \mathcal{N}(y|f(x,\theta),\sigma^2)$, то

$$\theta_{ML} = argmin_{\theta}\big[(y - X\theta)^2\big] = argmin_{\theta}\big[Loss(X,y,\theta)\big]$$

Maximum likelihood

Самым простым способом обучить такую модель будет найти максимум правдоподобия

$$\theta_{ML} = argmax_{\theta}\left[p(y|X,\theta)\right] = argmax_{\theta}\left[ln\left(p(y|X,\theta)\right)\right]$$

Как и говорили в прошлый раз, если $p(y|x,\theta) = \mathcal{N}(y|f(x,\theta),\sigma^2)$, то

$$\theta_{ML} = argmin_{\theta}\left[(y - X\theta)^2\right] = argmin_{\theta}\left[Loss(X,y,\theta)\right]$$

Можно показать, что $Loss(X,y,\theta)$ – выпуклая функция, а потому точка, в которой ее производная будет равняться нулю, будет точкой ее минимума

Closed form solution

Тогда

$$Loss = (y - X\theta)^\top (y - X\theta) =$$

$$= y^\top y - 2y^\top X\theta + \theta^\top X^\top X\theta$$

$$\frac{\partial Loss}{\partial \theta} = -2X^\top y + 2X^\top X\theta = 0$$

$$\theta = (X^\top X)^{-1} X^\top y$$

– closed form solution для линейной регрессии

Closed form solution

Проблемой closed form solution является:
1) $X^\top X$ чертовски экспансивная штука – её размерность $k \times k$, поэтому если k достаточно большое, считать обратную матрицу будет не простой задачей
2) Мы получили closed form solution только потому, что посчитать производную линейной функции и разрешить равенство её нулю относительно параметров – тривиальная задача, однако если бы функция была сложнее, то у нас не вышло бы сделать этого.

Gradient Descent

Альтернативой closed form solution является численная минимизация функции. Она основывается на том факте, что градиент функции – это вектор, направление которого совпадает с направлением возрастания функции, а потому мы можем выбрать некоторую начальную точку для параметров θ_0, после чего

$\theta_0 \leftarrow random\_number()$
for i in range(num_iterations) **do**
  $\theta_{i+1} \leftarrow \theta i - alpha * \frac{\partial Loss(X, y, \theta i)}{\partial \theta}$
end for

Stochastic Gradient Descent

Мы можем сделать алгоритм еще более эффективным – если мы выберем случайное подмножество данных $X_{mb} \subset X$ и аппроксимируем $\frac{\partial Loss(X, y, \theta i)}{\partial \theta}$ как $\frac{\partial Loss(X_{mb}, y_{mb}, \theta i)}{\partial \theta}$, то для вычисления производной нам понадобится существенно меньше вычислительных ресурсов

$\theta_0 \leftarrow random\_number()$
for i in range(num_iterations) **do**
  $(X_{mb}, y_{mb}) \leftarrow get\_random\_substet(X, y)$
  $\theta_{i+1} \leftarrow \theta i - alpha * \frac{\partial Loss(X_{mb}, y_{mb}, \theta i)}{\partial \theta}$
end for

Maximum Posteriori

Максимум правдоподобия может давать нам не самый лучший результат, когда мы захотим проверить работу нашего алгоритма на независимых данных.

Немного лучше в этом плане ведет себя Максимум Апостериора

$$\theta_{MAP} = argmax_{\theta}\Big[p(\theta|X, y)\Big] = argmax_{\theta}\Big[ln\Big(p(\theta|X, y)\Big)\Big]$$

$$= argmax_{\theta}\Big[ln\Big(p(y|X, \theta)\Big) + ln\Big(p(\theta)\Big)\Big]$$

Т.к. $p(\theta) \in (0; 1]$, то $ln\Big(p(\theta)\Big) \in (-\infty; 0]$, то мы можем сказать, что

$$Loss = (y - X\theta)^{\top}(y - X\theta) + f(\theta), f(\theta) \in [0; \infty)$$

Maximum Posteriori

Например

$$f(\theta) = \|theta\|_2$$

или

$$f(\theta) = \|theta\|_1$$

Bayes model averaging

$$p(\theta|X, y) = \frac{p(y|X, \theta)p(\theta)}{\int_\theta p(y|X, \theta)p(\theta)d\theta}$$

$$p(\hat{y}|X, y) = \int_\theta p(\hat{y}|X, \theta)p(\theta|X, y)d\theta$$