

Exercise 5. Task 2: Популяционные минимизаторы для разных потерь

Daniil Koveh

2025-11-06

Содержание

1	Теория	1
2	Жизненный пример	2
3	Академическое решение	2
3.1	Экспоненциальная потеря (AdaBoost)	2
3.2	ДевIANс (логистическая регрессия)	2
3.3	Квадратичная потеря	2
4	Проверка на числовом примере	2
5	Итог	3
6	Что запомнить	4

1. Теория

В задаче бинарной классификации Y
in

—1, 1 нам предлагают изучить три классические функции потерь:

1. **Экспоненциальная потеря** — в бустинге:

$$\text{ell}(f) = \exp(-Yf(x)).$$

2. **Логистическая (девианс)** — в логистической регрессии:

$$\text{ell}(p) = -\left[\frac{Y}{2} \log p(x) + \frac{1-Y}{2} \log(1-p(x)) \right].$$

3. **Квадратичная** — в регрессии:

$$\text{ell}(f) = (Y - f(x))^2.$$

Нужно найти **популяционные** минимизаторы, то есть аналитически вывести функцию $f^*(x)$ (или $p^*(x)$), которая минимизирует математическое ожидание потерь при фиксированном x .

2. Жизненный пример

Представим, что система определяет, совершит ли клиент покупку (1 — да, -1 — нет). Для каждого x (описание клиента) мы можем построить:

- **Boosting**: оптимальный логит $f^*(x)$, который потом превратится в вероятность.
- **Logistic regression**: сразу оптимальную вероятность $p^*(x)$.
- **Least squares**: оптимальную «регрессионную» оценку $f^*(x)$, которая затем переводится в вероятность.

Если понимать, как формулы появляются из математического ожидания, становится понятно, почему модели выдают именно такие значения.

3. Академическое решение

3.1. Экспоненциальная потеря (AdaBoost)

Ищем $f^*(x) = \arg \min_f \mathbb{E}_{Y|x} [e^{-Yf(x)}]$.

Разложим условное ожидание: $\mathbb{E}_{Y|x} [e^{-Yf}] = e^{-f} \Pr(Y = 1|x) + e^f \Pr(Y = -1|x)$. Дифференцируем по f и приравняем к нулю: $-e^{-f} \Pr(Y = 1|x) + e^f \Pr(Y = -1|x) = 0$. Отсюда $e^{2f} = \frac{\Pr(Y = 1|x)}{\Pr(Y = -1|x)}$. Берём логарифм: $f^*(x) = \frac{1}{2} \log \frac{\Pr(Y = 1|x)}{\Pr(Y = -1|x)}$. Это и есть логит условной вероятности. Симметрия ± 1 автоматически даёт коэффициент $1/2$.

3.2. Девианс (логистическая регрессия)

Используем параметризацию через вероятность $p(x) = \Pr(Y = 1|x)$. Потеря: $\ell(p) = -\left[\frac{Y+1}{2} \log p + \frac{1-Y}{2} \log (1-p) \right]$. Берём ожидание: $\mathbb{E}_{Y|x} [\ell(p)] = -\Pr(Y = 1|x) \log p - \Pr(Y = -1|x) \log (1-p)$. Задача: $p^*(x) = \arg \min_{0 < p < 1} \mathbb{E}_{Y|x} [\ell(p)]$. Это стандартная кросс-энтропия, минимум достигается при $p^*(x) = \Pr(Y = 1|x)$. Проверка: возьмём производную по p , приравняем к нулю и убедимся, что корень единственный и глобальный (функция строго выпуклая).

3.3. Квадратичная потеря

Потеря $(Y - f)^2$. Работает как в регрессии: $\mathbb{E}_{Y|x} [(Y - f)^2] = \mathbb{E}_{Y|x} [Y^2] - 2f \mathbb{E}_{Y|x} [Y] + f^2$. Минимум достигается в точке $f^*(x) = \mathbb{E}_{Y|x}(Y)$. Зная, что $Y \in \{-1, 1\}$ и обозначив $p = \Pr(Y = 1|x)$, получаем $\mathbb{E}_{Y|x}(Y) = 1 \cdot p + (-1) \cdot (1-p) = 2p - 1$. Значит, $f^*(x) = 2 \Pr(Y = 1|x) - 1$.

4. Проверка на числовом примере

```
p_true <- 0.3 # истинная вероятность класса +1
f_grid <- seq(-2, 2, length.out = 200) # сетка значений f
loss_exp <- p_true * exp(-f_grid) + (1 - p_true) * exp(f_grid) # экспоненциальная потеря
loss_sq <- p_true * (1 - f_grid)^2 + (1 - p_true) * (-1 - f_grid)^2 # квадратичная потеря
```

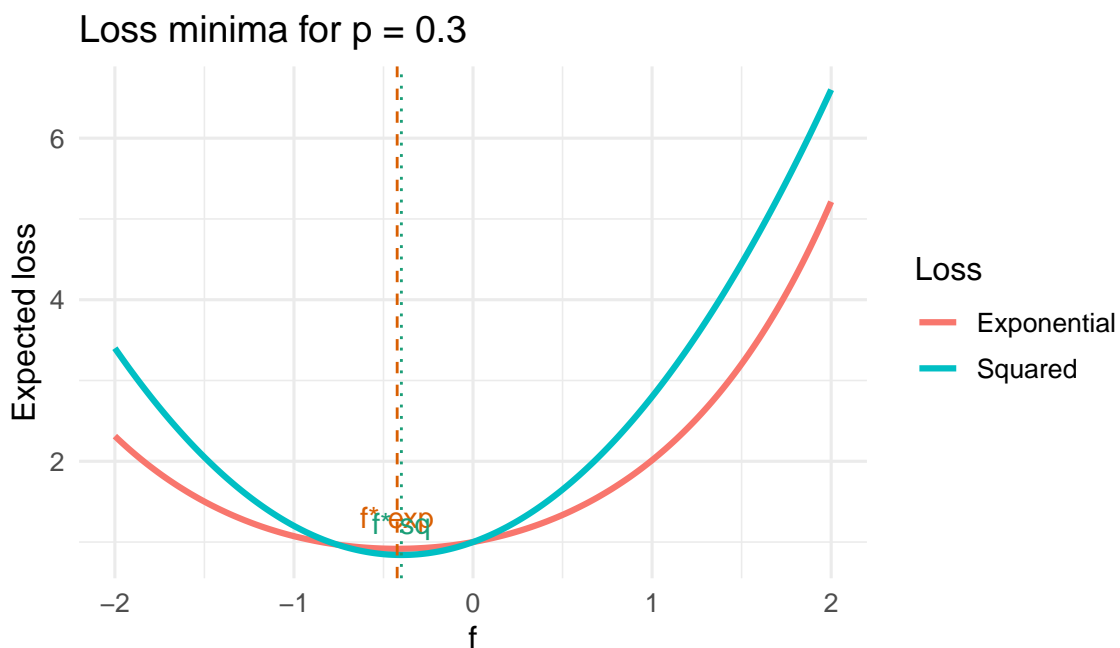
```

optimal_f_exp <- 0.5 * log(p_true / (1 - p_true)) # аналитический минимум
optimal_f_sq <- 2 * p_true - 1 # аналитический минимум

library(ggplot2) # графики
df_plot <- data.frame(
  f = rep(f_grid, times = 2),
  loss = c(loss_exp, loss_sq),
  loss_type = rep(c("Exponential", "Squared"), each = length(f_grid))
)

ggplot(df_plot, aes(x = f, y = loss, colour = loss_type)) +
  geom_line(size = 1.1) +
  geom_vline(xintercept = optimal_f_exp, linetype = "dashed", colour = "#d95f02") +
  geom_vline(xintercept = optimal_f_sq, linetype = "dotted", colour = "#1b9e77") +
  annotate("text", x = optimal_f_exp, y = min(loss_exp), label = "f* exp", vjust = -1, colour = "#d95f02") +
  annotate("text", x = optimal_f_sq, y = min(loss_sq), label = "f* sq", vjust = -1, colour = "#1b9e77") +
  labs(title = "Loss minima for p = 0.3",
       x = "f", y = "Expected loss", colour = "Loss") +
  theme_minimal(base_size = 12)

```



5. Итог

- Экспоненциальная потеря приводит к логиту условной вероятности: $f^*(x) = \frac{\log p}{\log(1-p)}$.
- Логистическая (девианс) напрямую возвращает вероятность: $p^*(x) = p$.
- Квадратичная потеря даёт условное среднее: $f^*(x) = 2p - 1$.

6. Что запомнить

- Все три потери минимизируются за счёт одного и того же объекта — условной вероятности $Pr(Y = 1|x)$.
- Разные потери просто переупаковывают вероятность по-разному: либо напрямую (логистическая), либо через логит (boosting), либо через центрирование на ± 1 (квадратичная).
- Понимание этих формул помогает интерпретировать выход любого алгоритма: когда модель возвращает логит, вероятность или среднее значение, мы точно знаем, как связать это с $Pr(Y = 1|x)$.