

Subsampling

- 데이터가 너무 크고 불균형인 문제를 해결한다.
- click이 일어나지 않은 경우를 $0 < r < 1$ 의 비율만 사용한다.\

- $$\frac{P(y = 1 | \mathbf{x})}{P(y = -1 | \mathbf{x})} = \frac{P(\mathbf{x} | y = 1)P(y = 1)}{P(\mathbf{x} | y = -1)P(y = -1)} = \frac{P'(\mathbf{x} | y = 1)P(y = 1)}{P'(\mathbf{x} | y = -1)P(y = -1)/r} = r \frac{P'(y = 1 | \mathbf{x})}{P'(y = -1 | \mathbf{x})}$$
- $$\log \frac{P'(y = 1 | \mathbf{x})}{P'(y = -1 | \mathbf{x})} = \mathbf{w}^T \mathbf{x} \Rightarrow \log r \frac{P'(y = 1 | \mathbf{x})}{P'(y = -1 | \mathbf{x})} = \log \frac{P(y = 1 | \mathbf{x})}{P(y = -1 | \mathbf{x})} = \mathbf{w}^T \mathbf{x} + \log r$$
- $\because P'(y = -1) = r \times P(y = 1) \text{ \& } P'(\mathbf{x} | y) = P(\mathbf{x} | y)$