Issues in Training Data

연습 문제 • 10 min

1. What formula represents a *dataset shift?* 1 / 1점

- $P_{train}(y|x) = P_{serve}(y|x)$ and $P_{train}(x) \neq P_{serve}(x)$
- $P_{train}(y, x) \neq P_{serve}(y, x)$
- $P_{train}(y|x) \neq P_{serve}(y|x)$ and $P_{train}(x) = P_{serve}(x)$

맞습니다

Well done! The most generic case of distribution skews is when the joint distribution of inputs and outputs differs between training and serving.

2. What measure is typically used to determine the degree of *data drift*? 1 / 1점

- Chebyshev distance (L-infinity)
- Euclidean distance (L2)
- Manhattan distance (L1)
- Hamming distance

맞습니다

That's it! Chebyshev distance is defined as $\max_i(|x_i - y_i|)$