

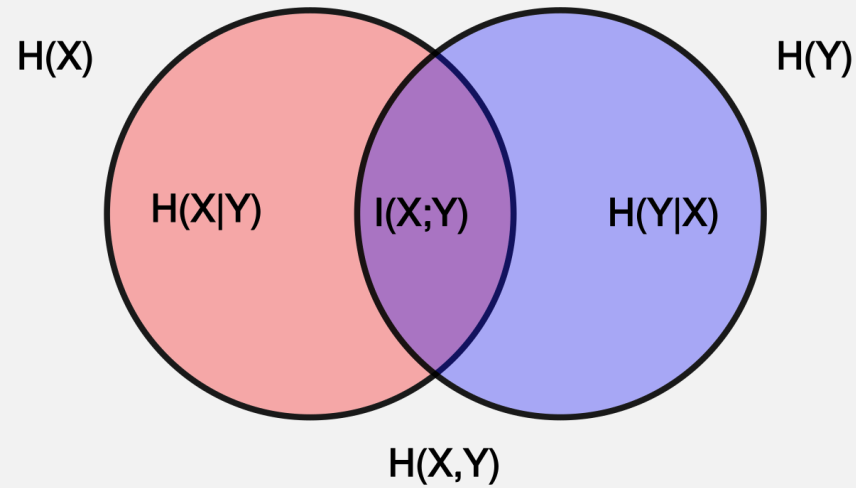


# Warsztaty modelowania

04 – feature engineering

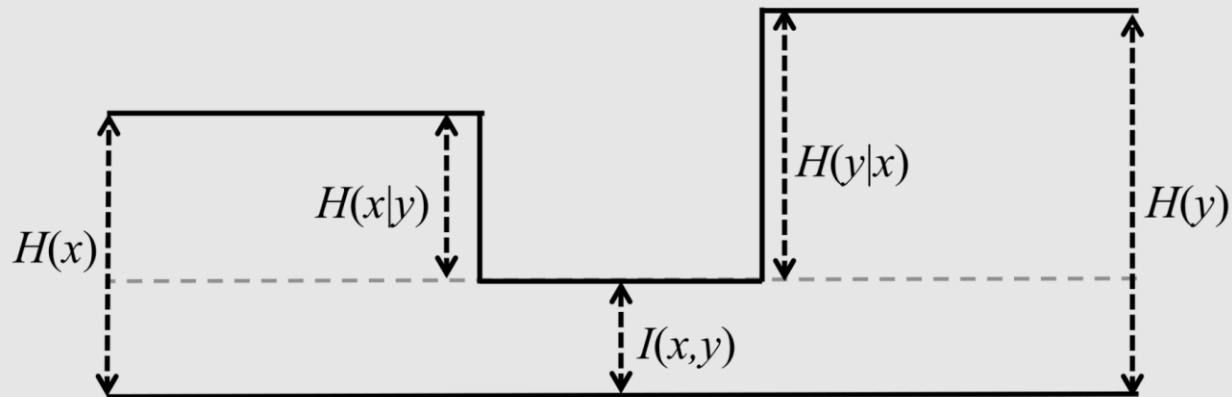
opracowała Patrycja Naumczyk

# Informacja wzajemna (mutual information)



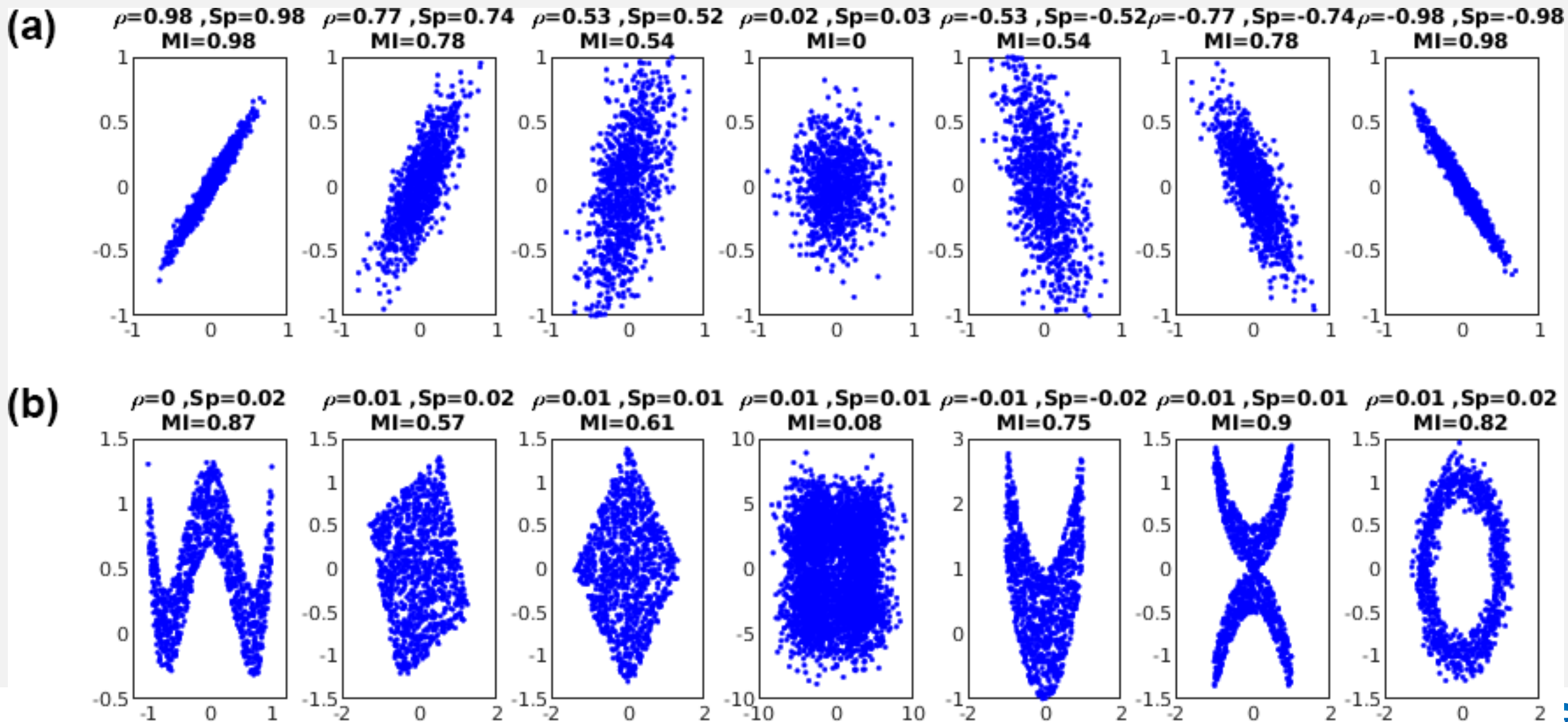
$$I(X; Y) = \sum_{y \in Y} \sum_{x \in X} p(x, y) \log \frac{p(x, y)}{p(x) p(y)}$$

$$I(X; Y) = \int_Y \int_X p(x, y) \log \frac{p(x, y)}{p(x) p(y)} dx dy$$



$$I(X; Y) = H(Y) - H(Y | X)$$

# Porównanie wartości informacji wzajemnej (MI) oraz korelacji Pearsona i Spearmana



# Informacja wzajemna w scikit learn

1. Ze sklearn.metrics – metryki pomiaru zgodności między etykietami rzeczywistymi, a przewidzianymi przez model:
  - a) mutual\_info\_score
  - b) adjusted\_mutual\_info\_score  
poprawka na entropię (mała liczba klas -> mała entropia)
  - c) normalized\_mutual\_info\_score  
znormalizowana 0-1, ale nie(!) korygowana o entropię
2. Z sklearn.feature\_selection – ocena przydatności zmiennej wyjaśniającej dla tłumaczenia zmiennej wyjaśnianej:
  - a) mutual\_info\_classif
  - b) mutual\_info\_regression

$$MI(U, V) = \sum_{i=1}^{|U|} \sum_{j=1}^{|V|} \frac{|U_i \cap V_j|}{N} \log \frac{N|U_i \cap V_j|}{|U_i||V_j|}$$

