

確認問題 (1)

1. 数値データをカテゴリデータに変換する必要があるのはどのような場合か考察せよ。また、このような変換の具体例を示せ。
2. カテゴリデータをニューラルネットワークの入力とする際に必要な処理について考えよ。
(ヒント：都道府県の情報はどう扱えばよい
か)

確認問題（１） 解答例

1. 別のデータと突き合わせることによって、個人が特定できてしまう場合がある。そのような場合は、数値をぼやかしてカテゴリで表現することによって、そのデータが表す個人を特定できないようにする。具体的には、数値データの値を特定の範囲で離散化する。たとえば年齢を表す数値データを 20 代、30 代 ... のようにカテゴリデータに変換する。

確認問題（１） 解答例

2. ニューラルネットワークは数値データを入力とすることが前提の学習アルゴリズムである。このような場合、0 または 1 の二値をとるダミー変数をカテゴリーの数だけ用意する。たとえば都道府県名を表すカテゴリデータは、47次元の one-hot ベクトル (1 つの次元の値だけが 1 で、残りは 0) となる。