第七章

EMアルゴリズム：最尤推定法による教師なし学習

EMアルゴリズムとは

期待値最大化法：

　反復法の一種であり、期待値ステップと最大化 ステップを交互に繰り替えすことで計算が進行する。

　Eステップでは、現在推定されている潜在変数の分布に基づいて、モデルの尤度の期待値を計算する。

　Mステップでは、E ステップで求まった尤度の期待値を最大化するようなパラメータを求める。

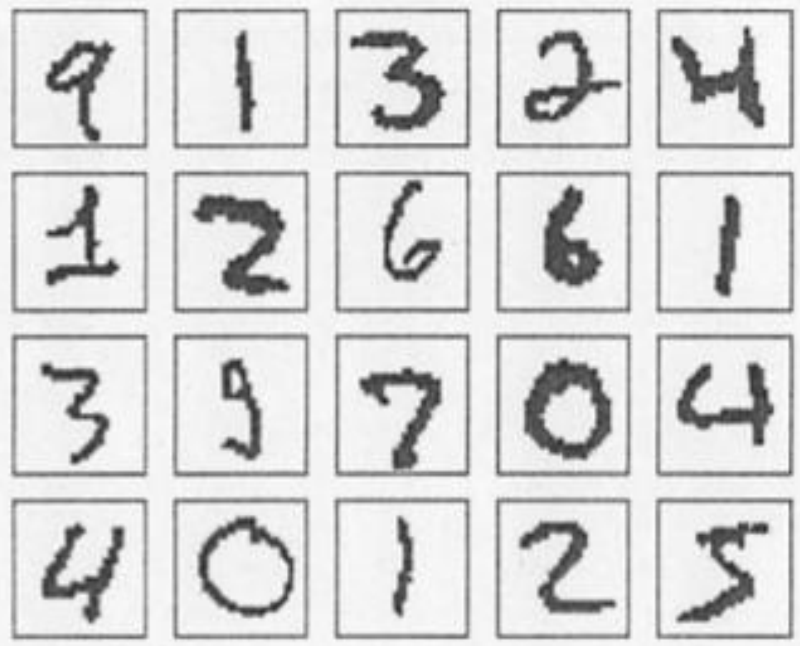
　M ステップで求まったパラメータは、次の E ステップで使われる潜在変数の分布を決定するために用いられる。

ベルヌーイ分布を用いた最尤推定法

本章で扱う例題は「手書き文字認識」

目標：手書きの数字の画像データが大量に与えられているとして、これらを分類すること。

方法：複数の顔写真を合成して、「平均的な顔」の写真を作る



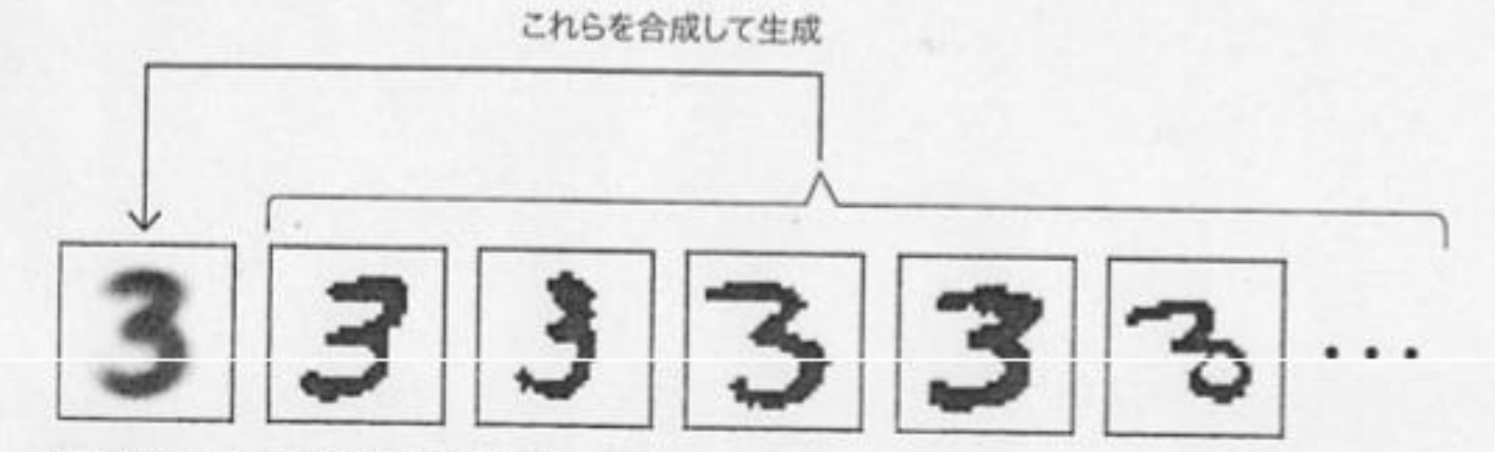
手書き数字の画像データ

手書き文字に合成方法

- 手書き文字の画像に含まれるピクセルを横一列にならべて、各ピクセルの色を1と0の数値で表したベクトルを用意。
- n番目の画像に対応するベクトルをx_nとし。第i成分$[x_n]_i$を見ると、i番目のピクセルの色がわかる。
- 特定の数字の手書き画像データがN個あるとして、これらの平均をとったベクトルμを用意。
- μ= $\frac{1}{N}\sum_{n=1}^{N} x_n$

- μの各成分は、0から1の範囲の実数値になる、これをピクセルの 色の濃淡と考え

- 実際にこの方法で100枚の手書き文字画像 を合成した結果です



▲図7.2 手書き文字画像を平均化した画像

画像生成器による最尤推定法の適用

- 最尤推定法を適用するには「トレーニングセットが得 られる確率」を計算する必要があり

- トレーニング セットに含まれる特定の画像データをxとし、xは、第i成分の値x_iが i番目のピクセルの色を表すベクトルである

- ピクセル数は、 全部でＮ個あるものとし、この中でも、特i番目のピクセ ルに注目した場合、そのピクセルの色が得られる確率p_iは次の式：

$x_i = 1$の場合：$p_i = \mu_i$
$x_i = 0$の場合：$p_i = 1 - \mu_i$

まとめて書くと
$p_i = {\mu_i}^{x_i}(1 - \mu_i)^{1-x_i}$

すべてのピクセルについて同じ色になる確率は次の式
$p(\mathrm{x}) = \prod_{i=1}^{D} p_i = \sum_{i=0}^{D} {\mu_i}^{x_i}(1 - \mu_i)^{1-x_i}$

すべてのデータ$[\mathrm{x_n}]_{\mathrm{n=1}}^{N}$を考える と、すべてに一致する画像が得られる確率は、次の式：
$\mathrm{P} = \prod_{n=1}^{N} \mathrm{p}(\mathrm{x_n}) = \prod_{n=1}^{N} \prod_{i=1}^{D} {\mu_i}^{[x_n]_i}(1 - \mu_i)^{1-[x_n]_i}$

このモデルにおける尤度関数になり

合分布を用いた最尤推定法

混合分布による確率の計算

全部でK個ある画像生成器を$[\mu_k]_{k=1}^{K}$とし、特定の画像生成器μ_kから画像xが得られる確率は次の式：

p_{μ_k}(x)= $\prod_{i=1}^{D}[\mu_k]_i^{x_i}(1-[\mu_k]_i)^{1-x_i}$

k番目の画像生成器を選ぶ確率をπ_kとして、$\{\pi_k\}_{k=1}^{K}$は次の条件を満たす

$\sum_{k=0}^{K}\pi_k$ =1

特定の画像xが得られる確率は次の式：

$$p(\mathrm{x}) = \sum_{k=1}^{K} \pi_k p_{\mu_k}(\mathrm{x})$$

最後に、トレーニングセットに含まれるデータ数をNとして、上記の操作をN回繰り返して、トレーニングセットのデータ群と一致する確率は次の式：

$$\mathrm{P} = \prod_{n=1}^{N} \mathrm{p}(\mathrm{x_n}) = \prod_{\mathrm{n}=1}^{N} \sum_{k=1}^{K} \pi_k p_{\mu_k}(x_n)$$

EMアルゴリズムの手続

- EMアルゴリズムというのは、このような形式の尤度関数を最大にするパラメーターを求める手続きになります

- K個の画像生成器$[\mu_k]_{k=1}^{K}$を適当に用意して、画像X“が得られる確率は、次の式：
- $p(x_n) \ = \sum_{k=1}^{K} \pi_k p_{\mu_k}(x_n)$

- 特定のk番目の画像生成器から、画像x_nが得ら れる可能性について、その割合を次式で取り出し
- $\gamma_{nk} = \frac{\pi_k p_{\mu_k}(x_n)}{\sum_{k=1}^{K} \pi_k p_{\mu_k}(x_n)}$

- 新たに画像生成器$[\mu_k]_{k=1}^{K}$を作り直して、それぞれの画像生成器を選択する確率$[\pi_k]_{k=1}^{K}$も計算し直す、次の計算式で、再設定しする：
- $\mu_k = \frac{\sum_{k=1}^{K} \gamma_{nk} x_n}{\sum_{k=1}^{K} \gamma_{nk}}$
- $\pi_k = \frac{\sum_{k=1}^{K} \gamma_{nk}}{N}$

この後は上記の二つの式決まった$[\mu_k]_{k=1}^{K}$と$[\pi_k]_{k=1}^{K}$を用いて、再度、γ_{nk}を計算しておいて、さらにまた、μ_kとπ_kを計算する、というようにこれらの手続きを何度も繰り返します。これを繰り返すごとに、Pの尤度関数の値が大きくなっていき、最終的に、極大値に達することが証明されています。