確率と最適化

―関数近似と回帰―

今回の目標

- 関数近似問題を勾配法を利用して解く

関数近似問題は、機械学習における「回帰の問題」に直接つながる。また、ニューラルネットの学習も一種の関数近似問題と捉えることができる。

勾配法

制約無し最適化問題

$$\text{minimize}_{\mathbf{x}\in\mathbb{R}^n} \ f(\mathbf{x}) \tag{1}$$

勾配法のステップ

Step 1 (初期点設定) $\mathbf{x} := \mathbf{x}_0$
Step 2 (勾配の計算) $\mathbf{g} := \nabla f(\mathbf{x})$
Step 3 (探索点更新) $\mathbf{x} := \mathbf{x} - \alpha\mathbf{g}$
Step 4 (反復) Step 2 に戻る

(注)

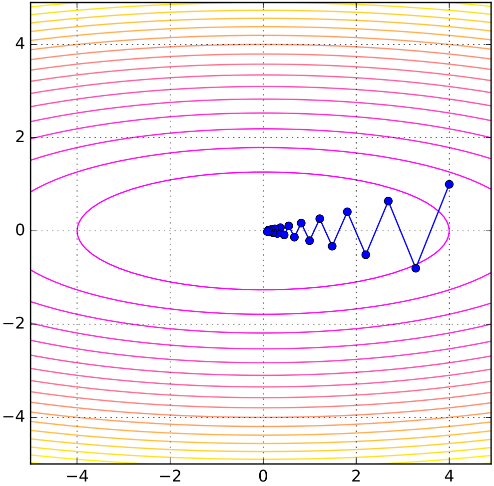
- α は探索係数 (実数)
- 探索点更新において直線探索を行う勾配法もある。

勾配法（雛形プログラム）

```
# -*- coding: utf-8 -*-
import numpy as np
# y = x^2 + y^2 を目的関数とする。
def grad(x):                          # 勾配
    return 2.0*x
xt = np.array([5,5])                # 初期点
alpha = 0.05                          # ステップ係数
for i in range(100):                  # メインループ
    xt = xt - alpha*grad(xt)
    print "step=",i,xt
    print
```

最小化の軌跡

$$f(x) = \frac{1}{2}x_0^2 + 5x_1^2, \quad \mathbf{x}_0 = (4, 1)^T, \alpha = 0.18$$



関数近似問題の問題設定

訓練集合: $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, $\{t_1, t_2, \ldots, t_N\}$

$$\mathbf{x}_i \in \mathbb{R}^n (i = 1, 2, \ldots, N), \quad t_i \in \mathbb{R}(i = 1, 2, \ldots, N)$$

データ点とそれに対応する観測値があるとき、データにフィットする関数を探したい！

関数近似問題のざっくりとした定義

与えられた訓練集合に対して、パラメータ付き関数 $f(\mathbf{x}|\mathbf{w})$ を利用して、すべての i で

$$t_i \simeq f(\mathbf{x}_i|\mathbf{w})$$

となるような $\mathbf{w}^*$ を求めたい。

回帰予測

訓練集合に含まれない未知の $\mathbf{x} \in \mathbb{R}^n$ がやってきたとき、

$$\hat{t} = f(\mathbf{x}|\mathbf{w}^*)$$

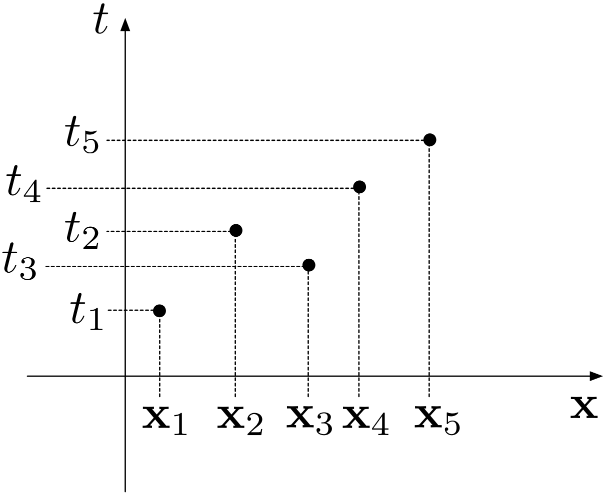
として予測を行う。

訓練集合

訓練集合: $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}, \quad \{t_1, t_2, \ldots, t_N\}$

$$\mathbf{x}_i \in \mathbb{R}(i = 1, 2, \ldots, N)$$

$$t_i \in \mathbb{R}(i = 1, 2, \ldots, N)$$

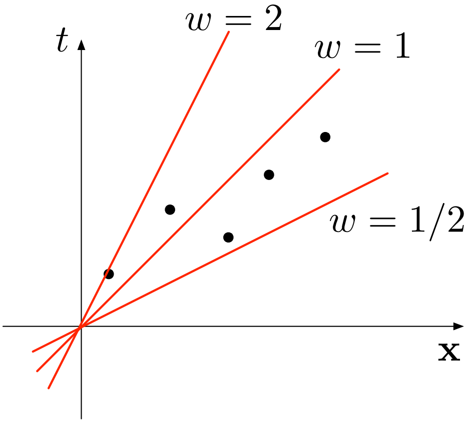


パラメータ付き関数の例

パラメータ付き関数の例として、

$$f(x|w) = wx$$

を考えよう。



もっともデータにフィットする w の選び方とは？

誤差関数

「関数近似問題のざっくりとした定義」$t_i \simeq f(\mathbf{x}_i|\mathbf{w})$ に含まれている $\simeq$ の意味を明確にしないと先に進めない。

2 乗和誤差関数

2 乗誤差を基準として

$$E(\mathbf{w}) = \frac{1}{2}\sum_{i=1}^{N}(f(\mathbf{x}_i|\mathbf{w}) - t_i)^2$$

と定義する。

2 乗和誤差関数に基づく関数近似問題

与えられた訓練集合 $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, $\{t_1, t_2, \ldots, t_N\}$ に対して、2 乗和誤差関数 $E(\mathbf{w})$ を最小化せよ。

練習問題

訓練集合として、$\{x_1 = 1, x_2 = 2\}, \{t_1 = 2.2, t_2 = 3.8\}$ が与えられた。パラメータ付き関数モデルとしては、$f(x|w) = wx$(傾き w の直線) を利用する。次の問いに答えよ。

(1) 2 乗和誤差関数 $E(w)$ を書け。
(2) 2 乗和誤差関数 $E(w)$ の勾配 (gradient) を求めよ。
(3) 勾配法のステップを実行してみよ。ただし、w の初期値は 0、$\alpha = 0.1$ とする。反復 3 回目までの計算を行え（電卓利用可）。

練習問題の解答

(1)

$$E(w) = \frac{1}{2}(wx_1 - t_1)^2 + \frac{1}{2}(wx_2 - t_2)^2 \tag{2}$$

$$= \frac{1}{2}(w - 2.2)^2 + \frac{1}{2}(2w - 3.8)^2 \tag{3}$$

(2)

$$\frac{\partial E(w)}{\partial w} = (w - 2.2) + 2(2w - 3.8) \tag{4}$$

$$= 5w - 9.8 \tag{5}$$

※普通だったら、ここで極値条件から直接 w を解くことになる ($E(w)$ は w に関して二次関数であり、下に凸であることが分かっているので「極値= 最小値」が保証される→ $w^* = 1.96$)。

練習問題の解答

(3) 問題より、初期値は $w_0 = 0.0$ であり、$\alpha = 0.1$ である。また、

$$g(w) = \frac{\partial E(w)}{\partial w} = 5w - 9.8$$

とおく。勾配法の反復計算過程は次のとおり。

(反復 1 回目) $g(0.0) = -9.8;\ w := 0.0 - 0.1(-9.8) = 0.98$

(反復 2 回目) $g(0.98) = -4.9;\ w := 0.98 - 0.1(-4.9) = 1.47$

(反復 3 回目) $g(1.47) = -2.45;\ w := 1.47 - 0.1(-2.45) = 1.715$

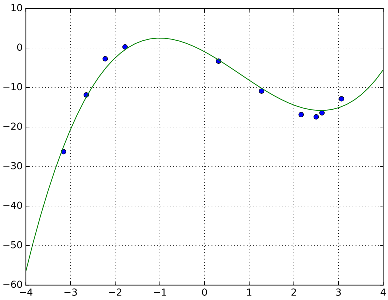
パラメータ付き関数 $f(\mathbf{x}|\mathbf{w})$ の選択 (1)

<u>多項式モデル</u>

$x \in \mathbb{R}$

$$f(x|\mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M$$

M をこの多項式モデルの次数と呼ぶ。



3 次関数モデルによるフィッティング

パラメータ付き関数 $f(\mathbf{x}|\mathbf{w})$ の選択 (2)

<u>正弦波重ね合わせモデル</u>

$$f(x|\mathbf{w}) = w_0 + w_1 \sin(\omega x) + w_2 \sin(2\omega x) + \cdots + w_M \sin(M\omega x),\ x \in \mathbb{R}$$

(ω は与えられた定数)

<u>線形回帰モデル</u>

$$f(\mathbf{x}|\mathbf{w}) = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_M x_M, \quad x = (x_1, \ldots, x_M) \in \mathbb{R}^M$$

練習問題

線形回帰モデル (w_0 としている)

$$f(\mathbf{x}|\mathbf{w}) = w_1 x_1 + w_2 x_2$$

を考える。訓練集合は

$$\begin{aligned} \mathbf{x}_1 &= (1, 2),\ t_1 = 4 \qquad (6) \\ \mathbf{x}_2 &= (3, 1),\ t_1 = 2 \qquad (7) \end{aligned}$$

と与えられている。次の問いに答えよ。

(1) 誤差関数のヘッセ行列が正定値であることを示せ。

(2) 解析的に w^* を求めよ (すなわち、勾配法は利用しない)。

練習問題解答

(1) 誤差関数 $E(w_1, w_2)$ は

$$E(w_1, w_2) = \frac{1}{2}(w_1 + 2w_2 - 4)^2 + \frac{1}{2}(3w_1 + w_2 - 2)^2$$

と与えられる。この関数を w_1, w_2 でそれぞれ偏微分すると

$$\frac{\partial E}{\partial w_1} = 10w_1 + 5w_2 - 10 \qquad (8)$$

$$\frac{\partial E}{\partial w_2} = 5w_1 + 5w_2 - 10 \qquad (9)$$

を得る。さらにヘッセ行列は、

$$\begin{pmatrix} 10 & 5 \\ 5 & 5 \end{pmatrix}$$

となる。

練習問題解答

2 次形式

$$
\begin{aligned}
(a,b)\begin{pmatrix} 10 & 5 \\ 5 & 5 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} &= 10a^2 + 10ab + 5b^2 && (10) \\
&= 5a^2 + 5(a+b)^2 && (11)
\end{aligned}
$$

は任意のゼロではない $a, b \in \mathbb{R}$ に対して、正の値を持つ。したがって、このヘッセ行列は正定値行列であり、$E(w_1, w_2)$ は下に凸 (convex) である。

(2) この誤差関数は下に凸な関数であることがわかったので、極値条件を満たす点が最小値を与える。極値条件 $\nabla E(w_1, w_2) = 0$

$$
\begin{aligned}
\frac{\partial E}{\partial w_1} &= 10w_1 + 5w_2 - 10 = 0 && (12) \\
\frac{\partial E}{\partial w_2} &= 5w_1 + 5w_2 - 10 = 0 && (13)
\end{aligned}
$$

を解くことにより $w_1^* = 0, w_2^* = 2$ を得る。

回帰フレームワークはどんなところで有用か？

例えば、興味のある対象について、過去のデータから将来の観測値を予測する、という使い方ある。

例 1: 株価予測

訓練集合 第 i 日の円ドル為替レート、第 i 日の X 社の株価、そのほか経済指標
$\rightarrow$

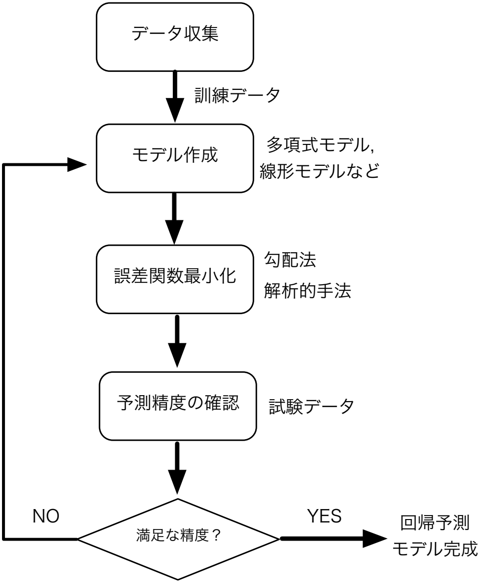
予測目標 将来ある時点での X 社の株価

例 2: 自動車事故率 (保険会社の視点から)

訓練集合 契約者の年齢、過去の事故歴、車種$\rightarrow$

予測目標 将来のある人の事故率

回帰のためのレシピ


データ収集
訓練データ
モデル作成
多項式モデル,
線形モデルなど
誤差関数最小化
勾配法
解析的手法
予測精度の確認
試験データ
満足な精度？
NO
YES
回帰予測
モデル完成