

1 SVM, Gaussian kernel で training した

training set には よく fit して いる。 } → overfitting
 cross validation set には fit しない

$C = \frac{1}{\lambda}$ $\lambda \rightarrow$ 大ならば " $\theta_1 \sim \theta_n$ を 小さく しやすいので、相対的に $\theta_0 \rightarrow$ 大 \therefore high Bias
 ($C \rightarrow$ 小) 正しい選択

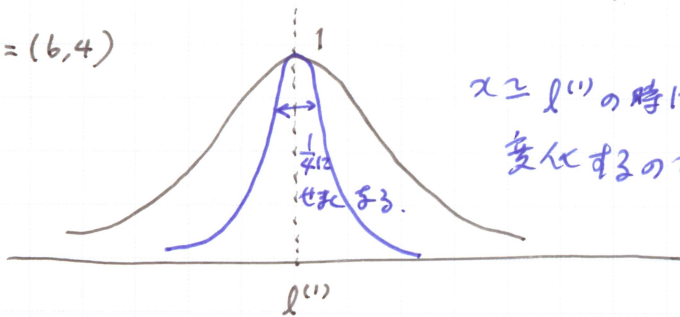
$\sigma^2 \rightarrow$ 大 $f_i = \exp\left(-\frac{\|x^{(i)} - l^{(j)}\|^2}{2\sigma^2}\right)$ $x^{(i)} = l^{(j)}$ の時の 変化が ゆるやか になり \therefore high Bias

underfitting 方向に、即ち high Bias に 変化させたいので
 $C \rightarrow$ 小, $\sigma^2 \rightarrow$ 大

2 Gaussian kernel を使った similarity $(x, l^{(1)}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)$

図は $\sigma^2 = 1$ の時。 $\sigma^2 = 0.25 = \frac{1}{4}$ に したと いう 事 がある。

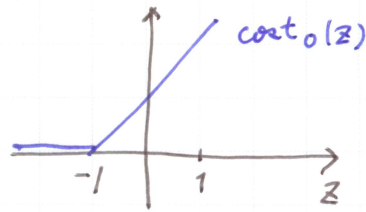
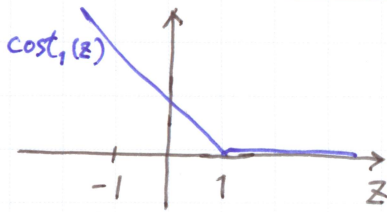
$l^{(1)} = (6, 4)$



$x = l^{(1)}$ の時に e^{-2} で z の 値 が 4 倍 速く 変化 する ので。

Week 7 Quiz 2

$$\min_{\theta} C \sum_{i=1}^m y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) + \sum_{j=1}^n \theta_j^2$$



$y^{(i)} = 1$ のとき $\theta^T x \geq 1$

$y^{(i)} = 0$ のとき $\theta^T x \leq -1$

□ の部分が 0 になる条件を 2つ 選択する ↑ この2つを 選ぶ

✗ 4 $n = 10$ features , $m = 5000$ examples

gradient descent で logistic regression を学習 (train) した。

training set \rightarrow underfit

cross validation set \rightarrow =

- ① 高次の feature を追加する。 overfit 方向 への \rightarrow
- ② training set の数を減らす。 ✗ 論外
- ③ 最適化手法を変更する。 \because logistic regression を学習する gradient descent が local minimum にハマっているの \therefore ○ おまり fit していないのは local minimum の状態
- ④ たくしんの hidden unit を持つ neural network を使う
学習させるのはたいてい OK だが、やはり overfit はできる。 ○ とした。 微妙なところ...

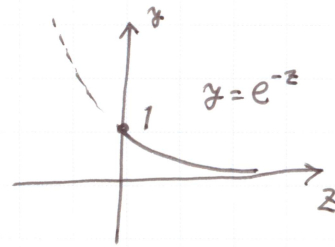
5

① multi-class classification に SVM を用いて one-vs-all アプローチを使う。

K 個の異なるクラスには (K-1) 個の SVM が必要。 ~~X~~ K 個必要
 クラス i と「それ以外」を分類する必要があるため。

② Gaussian Kernel を使う前に feature normalization を実行しておくことが重要である。 そうしないと特定の feature に影響されてしまう。

③ Gaussian kernel の最大値は 1 である。
 $y = e^{-z}$, $z \geq 0$



④ データが linear に分離できる時、

linear kernel を用いた SVM は C の値によらずに同じ θ を返す

$$\min_{\theta} \left(\sum_{i=1}^m y^{(i)} \text{cost}_1(\theta^T x^{(i)}) + (1 - y^{(i)}) \text{cost}_0(\theta^T x^{(i)}) \right) + \sum_{j=1}^n \theta_j^2$$

→ 0 になる。 ∴ C によらない

$y^{(i)} = 1$ のとき全ての i について
 $\theta^T x^{(i)} \geq 1$

$y^{(i)} = 0$ の時全ての i について
 $\theta^T x^{(i)} \leq -1$

これが可能

1 Gaussian kernel の SVM → underfitting

$C = \frac{1}{\lambda}$ $\lambda \rightarrow$ 小なると θ_1, θ_n を小さくする影響は小。相対的に $\theta_0 \rightarrow$ 小 \therefore high variance
 (C → 大) 正しい overfitting

$\alpha^2 \rightarrow$ 小 $x = l^{(i)}$ のまわりで変化が急。 \therefore overfitting

2 $\alpha^2 = 1$ のグラフが示された。 $l^{(1)} = (6, 4)$
 $\alpha^2 = 0.25$ にしたらどうなる。山の高さは1のままに幅がせまくなる。

3 $y^{(i)} = 1$ のとき $\theta^T x^{(i)} \geq 1$
 $y^{(i)} = 0$ のとき $\theta^T x^{(i)} \leq -1$ } を選ぶ

X 4 $n = 10$ feature, $m = 5000$ examples
 gradient descend を用いた logistic regression → training set
 cross validation set } \therefore underfit

① 別の optimization method を使え。local minimum に陥っている ○

② training set の数を減らせ X 論外

③ たくさん hidden unit を持つ neural network を使え X \therefore 足りない
 \therefore hidden unit は必要ない

④ 高次の feature を追加する ○

X [5]

① multi-class SVM を one-vs-all で学習する時, kernel を使う必要はない
X 論外

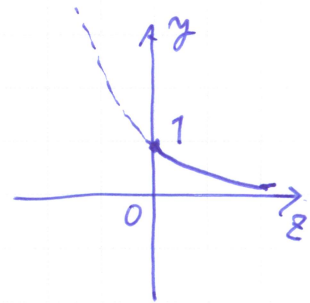
② データが linear に分離できる時, linear kernel を使う SVM は C の値によらずに同じ θ を返す X

$$\min_{\theta} C \left[\sum_i \right] + \left[\sum_{j=1}^n \theta_j^2 \right]$$

線形分離 \uparrow これを 0 にすることはできる。しかし、その時 C が大きくなる、という意味がある。
(0 にできること)

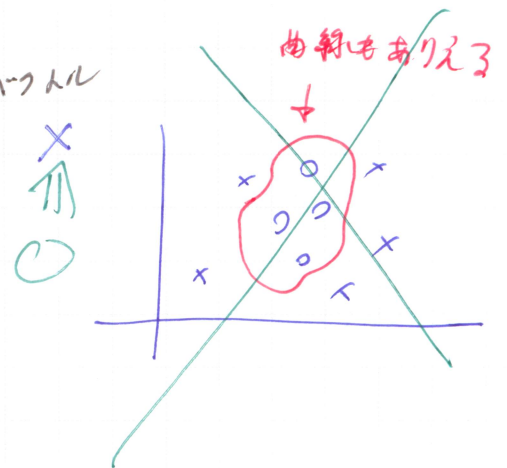
あくまでも $C \square + \square$ を最小化するのだから。
 $\therefore C$ の値によって θ は変化する。

③ Gaussian kernel の最大値は 1 である ○



④ $x^{(i)} \in \mathbb{R}^2$ かつ $x^{(i)} = (x_1, x_2)$ 2次元ベクトル
SVM の decision boundary は直線である X

linear kernel の
読み落としてはいけないこと



Week 7 Quiz 3回目

1 Gaussian kernel の SVM の training set に underfit
overfit 方向にしたい

$C = \frac{1}{\lambda}$ $\lambda \rightarrow \text{小}$ overfit 方向

$\therefore C \rightarrow \text{大}$

$\alpha^2 \rightarrow \text{小}$ overfit 方向

2 $\alpha^2 = 1$ の時 $\alpha^2 = 0.25$, せよくなる

3 $\gamma = 1$ の時 $\phi^T x^{(i)} \geq 1$, $\gamma = 0$ の時 $\phi^T x^{(i)} \leq -1$

X 4 $n=10$, $m=5000$, logistic regression classifier with gradient descent underfit

- ① training set の数を減らせば X
- ② 多くの hidden unit を持つ neural network を試せば X ← 0
- ③ 高次の feature を追加した X ← 0
- ④ local minimum におさまるの optimization 手法をかえれば O ← 20000 x 20000 ? local minimum へ近づける

5 ① multi-class SVM with one-vs-all approach , kernel は使わない X

② 線形分離できないデータは C の値をよすすに , 同じ ϕ が求まる X

C の項目は ϕ に依存する $\sum_{j=1}^n \phi_j^2$ とあわせて
 最小にするように ϕ , C の項目を 0 に近づける
 ことをあつた

③ Gaussian kernel の最大値は 1 である O

④ 2次元の feature に対して linear kernel の SVM の decision boundary は直線 O

前回も読みかかっていた

Week 7 Quiz 4回目と5回目

- ✗ [4] ① neural network with a large number of hidden unit ✗
 local minimum あり
 ② 異なる optimization method を使う ✗
 ③ training set の数 を 減らす ✗
 ④ polynomial feature を 追加 した ○

[4] $n=10, m=5000$, gradient descent \rightarrow underfit
 正解

- ① linear kernel の SVM を 使う. (feature を 追加 して やる) ✗
 ② polynomial feature を 追加 した ○
 ③ Gaussian kernel の SVM を 使う ○
 ④ λ の 値 を 増やせ ✗

ポイント は underfit を防ぐ. overfit 方向 に 動かす こと. feature 追加 は その ため.
 gradient descent の local minimum は, その 場合 にも 対処 が必要. 断言 した ✗

$n=10, m=5000$ かつ
 n 小 $1 \sim 1000$ m 大 $10 \sim 10000$
 } SVM with Gaussian kernel かなり 適切.
 } neural network $\leftarrow n, m$ に よる が 適切

(注) n 大 10000 , m 小 $10 \sim 1000$ では { logistic regression
 linear kernel の SVM
 n 小 $1 \sim 1000$, m 大 50000 以上 ならば { :

[4] ① local minimum ありで, 別の optimization method を 使う. \rightarrow いや, 初期値 を 変え たり して
 gradient descent の optimization method は 対応 する
 べし

✗ が 正しい こと を
 確認 した