Poly-Regression-Bias-Variance

Monday, 9 August 2021 7:13 PM

Agenda :

- Polynomial Regression
- Generalization and Occam's Razor
- Underfit Vs Overfit
- Bias Variance Tradeoff
- Regularization
- L1 Vs L2 Vs Elastic net
- Hyperparameter

Polynomial Regression

linear reg ✓

linear reg + high [illegible] (x^2)

linear reg $\hat{y} = w_0 + w_1 f_1$

and $f_2 = f_1^2$ $\Rightarrow$ $\hat{y} = w_0 + w_1 f_1 + w_2 f_1^2$

$y^{(i)} \leftarrow f_{new} \left(f_1, f_1^2 \right)$

"Polynomial"

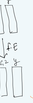
$f_2 = \alpha f_1 + \beta$

"Spanning" in new feature.

No ML

trained my model on

f_1, f_1^2



Scatter plot

$\Rightarrow$

Scatterplot (f_1, y)

modification (f_1, f_1^2)

$$y = 0.7x^5 + 2.1x^4 + 0.37x^3 + 0.2x^2 + 0.3x$$

$\Rightarrow$

f_1 deg = 1

f_2, f_3 deg = 2

$1, f_1, f_2, f_1^2$

$1, x, x^2$

Generalization & Occam's Razor

(lower degree Polynomial)

only f_1

M1

(Higher degree Polynomial)

$f_1, f_1^2, f_1^3, f_1^4, \ldots$

M2

Q → which is better ?

$$MSE \Rightarrow \frac{1}{n} \sum_i \left(y^{(i)} - \hat{y}^{(i)} \right)^2$$

① Generalization

D → D_train, D_test, D_Unseen

; check the performance.

② Occam's Razor : [Science → ML]

Problem → $S_1, S_2, S_3, \ldots S_{10}$

$S_9, S_{10} \rightarrow S_1, S_2, S_3, S_4, S_5, S_6, S_7, S_8$

$\Rightarrow$ 2 K the simplest S_1^n

① linear : f_1

② Quad : f_1, f_1^2

③ Cubic : f_1, f_1^2, f_1^3

Simplest (linear) M1 — H

M2 (w_0, w_1, w_2) — L

Complex (w_0, w_1, w_2, w_3) M3 — L

Generalization Test-score

$M_1 \rightarrow f_1$

$M_2 \rightarrow f_1, f_1^2$

$M_{10} \rightarrow f_1, f_1^2 \ldots f_1^{10}$

(w_0, w_1) M1 — Underfit

(w_0, w_1, w_2) M2 — Perfectly just-right

$(w_0, w_1, w_2, \ldots w_{10})$ M10 — overfit

	Train	Test
Underfit	100	120
Perfect	20	30
Overfit	5	50

Simplest → complex

Large

Mathematical Intuition

$$y = w_0 + w_1 f_1 + w_2 f_1^2 + w_3 f_1^3 + w_4 f_1^4$$

transformed features

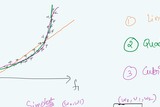
originally → f_1

① $w_1, w_2, w_3, w_4 \neq 0$ — overfit

② $w_3 = w_4 = 0$, $w_1, w_2 \neq 0$ — weight

③ $w_2 = w_3 = w_4 = 0$, $w_1 \neq 0$ — underfit

Bias Variance Tradeoff



Error $\Rightarrow$ Bias^2 + Variance + irreducible error

$f_1, f_2 \ldots f_5$ — f_{25}

① increase the complexity

create f_{25} / $f_{\frac{1}{2}}$

other nonlinear features

VIF → ∞

None

$VIF_i = \frac{1}{1 - R_i^2}$

$R_i \rightarrow \infty$

No linear Regression

$f_2 = \alpha_1 f_1 + \alpha_2 f_2 \ldots + \alpha_n f_n$

$f_2 \leftarrow (f_1, f_3 \ldots f_n)$

If VIF > 5

VIF > 10

VIF > 2

VIF > 1

VIF > 0.1