Midterm Cheat Sheet

Linear Regression

Given an input $X = \{x_1, \ldots, x_n\}$, and output y,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

where β_0 is intercept, and $B = \{\beta_0, \beta_1, \ldots, \beta_n\}$ are model coefficients.

Polynomial Regression

Given an input x, output y, and degree d

$$y = \beta_0 + \beta_1 x^1 + \cdots + \beta_n x^n$$

where β_0 is intercept, and $B = \{\beta_0, \beta_1, \ldots, \beta_n\}$ are model coefficients.

Logical Regression

Given an input $X = \{x_1, \ldots, x_n\}$, weights $W = \{w_0, w_1, \ldots, w_n\}$, and threshold t,

$$z = w_0 + w_1 x_1 + \cdots + w_n x_n$$

$$y = \begin{cases} 1 & \text{if Sigmoid}(z) >= t \\ 0 & \text{if Sigmoid}(z) < t \end{cases}$$

Regularization

$$\sum_{i=1}^{M} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{M} (y_i - \sum_{j=0}^{N} w_j \cdot x_{ij})^2 + R$$

where, R is the regularization term.

L1 Regularization

$$R_{L1} = \lambda \sum_{j=0}^{N} |w_j|$$

L2 Regularization

$$R_{L2} = \lambda \sum_{j=0}^{N} w_j^2$$

Midterm Cheat Sheet

Multilayer Perceptron Model

Given an input $X = \{x_1, \ldots, x_n\}$, bias $x_0 = 1$, and weights $W = \{w_0, w_1, \ldots, w_n\}$,

$$
\begin{aligned}
z &= w_0 + w_1 x_1 + \cdots + w_n x_n \\
y &= \text{Activation}(z)
\end{aligned}
$$

Activation Functions :

- $\text{BinaryStep}(z) = \begin{cases} 1 & \text{if } z >= 0 \\ 0 & \text{if } z < 0 \end{cases}$
- $\text{Sigmoid}(z) = \dfrac{1}{1 + e^{-z}}$
- $\text{ReLU}(z) = \begin{cases} z & \text{if } z >= 0 \\ 0 & \text{if } z < 0 \end{cases}$
- $\text{LeakyReLU}(z) = \begin{cases} z & \text{if } z >= 0 \\ 0.01z & \text{if } z < 0 \end{cases}$
- $\tanh(z) = \dfrac{e^{2z} - 1}{e^{2z} + 1}$

Residual Sum of Squares (RSS)

$$
\begin{aligned}
RSS &= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \\
&= \sum_{i=1}^{n} (y_i - w^T x_i)^2
\end{aligned}
$$

Ordinary Least Squares (OLS) method

$$
w \triangleq argmin_w RSS = argmin_w \left(\sum_{i=1}^{m} (y_i - w^T x_i)^2\right)
$$

Gradient Descent Method

Given w_j is current rate, $\frac{\delta RSS}{\delta w_j}$ is gradient, and α is learning rate, new weight w_j' is given by,

$$
w_j' = w_j - \alpha \frac{\delta RSS}{\delta w_j}
$$

1. False. there is no x_1^2 and x_1^3 terms
not all input's nth degree terms are present.

2. b

3. c.

$$\frac{1}{1+e^{4.5}} = 0.01 \rightarrow \text{sigmoid}$$

relu: 0

4. N_1 : NAND
N_2 : OR
overall: XNOR

4. cont.

X_1	X_2	N_1	N_2	N_3
0	0	1	0	1
0	1	1	1	0
1	0	1	1	0
1	1	0	1	1

only 0 if $x_1 = x_2 = 1$

$1 - 0.5 \times 2 = 0$

≥ 0.5 o.w.

$N_1 = 1 - 0.5X_1 - 0.5X_2$

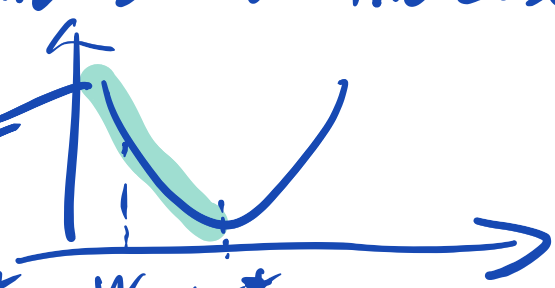
only 0 if $x_1 = x_2 = 0$

$\because -0.5 < 0.5$

$N_2 = -0.5 + 1.5X_1 + 1.5X_2$

$N_3 = 1.5 - N_1 - N_2 \Rightarrow$ only 0 if $N_1 = N_2 = 1$

5. weight should increase



6. 0.1567

$y_1 = -5, \hat{y}_1 = 0.5 + 0.8 \times 1 - 0.6 \times 6 - 3$

$= -5.3$

$y_2 = 0.4, \hat{y}_2 = 0.5 + 0.8 \times 2 - 0.6 \times 4 + 0.5$

$= 0.2$

$y_3 = 6.3, \hat{y}_3 = 0.5 + 0.8 \times 8 - 0.6 \times 2 + 1.5$

$= 7.2$

$$\frac{1}{2}\left(\frac{(5.3-5)^2 + (0.4-0.2)^2 + (7.2-6.3)^2}{3}\right)$$

≈ 0.1567

7. b.

8. d

9. b.

10. classify as 0, 0, 1, respectively

$z_1 = 0.3\times1.5+0.4\times4-0.9\times6$

$= -3.35$

$\phi(z_1) = 0.0338 \Rightarrow$ classify as 0

$z_2 = 0.3\times7+0.4\times3.5-0.9\times5$

$= -1$

$\phi(z_2) = 0.2689 \Rightarrow$ classify as 0

$z_3 = 0.3\times2+0.4\times8-0.9\times1$

$= 2.9$

$\phi(z_3) = 0.9478 \rightarrow$ classify as 1

11. c

12. c

13. b.

$0.8 + 0.5x_1 + 0.5x_2$

as long as one x_i is 1

$y = 1$ since $0.8 + 0.5 \geq 1$

So OR gate

14. $w^{(t+1)} = Bw^{(t)}$, $B = \begin{bmatrix} 3/4 & 0 \\ 0 & 1/2 \end{bmatrix}$

$$J(w) = \frac{1}{2} w^T A w$$

$$\frac{\partial J(w)}{\partial w} = \frac{1}{2} \frac{\partial (w^T A w)}{\partial w}$$

$$= \frac{1}{2}(2Aw)$$

$$= Aw$$

$$\therefore \Delta J(w^{(t)}) = Aw^{(t)}$$

$$w^{(t+1)} = w^{(t)} - \frac{1}{4} \nabla J(w^{(t)})$$

$\frac{1}{4}A = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/2 \end{bmatrix}$

$$= w^{(t)} - \frac{1}{4} A w^{(t)}$$

$$= (I - \frac{1}{4}A) w^{(t)}$$

$$= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 1/4 & 0 \\ 0 & 1/2 \end{bmatrix} \right) w^{(t)}$$

$$= \begin{bmatrix} 3/4 & 0 \\ 0 & 1/2 \end{bmatrix} w^{(t)}$$

$$= B w^{(t)}$$

15.

no regularization term in 15.1

15.1 $\min \sum (y_i - w^T x_i)^2$

15.2 $\min \sum (y_i - w^T x_i)^2 + \lambda \sum_{j=1} w_j^2$

15.3 $\min \sum (y_i - w^T x_i)^2 + \lambda \sum_{j=1} |w_j|$

For large λ in 15.2

(c) because w_0 (bias) is unaffected by regularization terms

16.

(1) $\vec{e} = \underset{n\times 1}{y} - \underset{n\times k}{X}\underset{k\times 1}{\hat{\beta}}$

(2) $\sum_i e_i^2 = \vec{e}^T\vec{e} = (y - X\hat{\beta})^T(y - X\hat{\beta})$

(3) $\vec{e}^T\vec{e} = (y - X\hat{\beta})^T(y - X\hat{\beta})$

$= (y^T - \hat{\beta}^T X^T)(y - X\hat{\beta})$

$= y^Ty - y^TX\hat{\beta} - \hat{\beta}^TX^Ty + \hat{\beta}^TX^TX\hat{\beta}$

$\frac{\partial(\vec{e}^T\vec{e})}{\partial \beta^*} = 0$

$\text{scalar}^T = \text{scalar}$

$= (y^TX\hat{\beta})^T = y^TX\hat{\beta}$

$\partial(y^Ty - y^TX\hat{\beta} - \hat{\beta}^TX^Ty + \hat{\beta}^TX^TX\hat{\beta}) / \partial\beta^* = 0$

$\Rightarrow \quad \frac{\partial(-2y^TX\hat{\beta})}{\partial\hat{\beta}^*} + 2X^TX\hat{\beta} = 0$

$-2y^TX + 2X^TX\hat{\beta} = 0$

$\beta^* = (X^TX)^{-1}X^Ty$