

Linear Regression

Gradient Descent

- **Model** $y = w \cdot x + b$
 - Model \rightarrow **training data** \rightarrow **testing data** \rightarrow **Overfitting**
- **Loss** $L(w, b) = \sum_{i=1}^n (y_i - (w \cdot x_i + b))^2$
- **Minimization**
 - $w_1 = w_0 - \alpha \frac{\partial L}{\partial w}|_{w=w_0}$
 - $b_1 = b_0 - \alpha \frac{\partial L}{\partial b}|_{b=b_0}$
 - α **Learning rate**

Regularization

- **Loss function**
 - $L(w, b) = \sum_{i=1}^n (y_i - (w \cdot x_i + b))^2 + \lambda \sum (w_i)^2$
 - λ **Loss function**
 - λ \rightarrow 0 \rightarrow **Underfitting**
 - λ \rightarrow **Overfitting**

Classification

- **1. 機械学習**
 - **条件付き確率**
 - $P(y=k|x)$ \rightarrow **条件付き確率**
 - **全確率の定理**
- $$P(y=k|x) = \frac{P(x | y = k) P(y = k)}{P(x)}$$
- ここで
- $P(y=k|x) \propto P(x | y = k) \propto P(y = k)$
 - $P(x | y=k) \propto P(y=k | x) \propto P(x)$
 - $P(y=k) \propto 1$

-
- **2. ロジスティック回帰**
 - $P(y=k|x)$ \rightarrow $P(x|y=k)$ \rightarrow $P(y=k)$ \rightarrow **条件付き確率**
 - **条件付き確率**
 - **条件付き確率**

-
- **3. ロジスティック**
 1. $P(y=k)$
 2. $P(x|y=k)$
 - $x \cdot P(y=k)$

3. ** $P(y=k|x)$ **

4. မြန်မာစာ

- မြန်မာစာ **MLE** မြန်မာစာ μ Σ မြန်မာစာ

 - $L(\mu, \Sigma) = f_{\mu, \Sigma}(x^1)f_{\mu, \Sigma}(x^2)\dots f_{\mu, \Sigma}(x^{79})$
 - $L(\mu, \Sigma) \propto \mu^{-79} \exp(-\frac{1}{2}\sum_{i=1}^{79} (x_i - \mu)^2)$
 - $\mu \sim \mathcal{N}(\mu_1, \Sigma)$

- မြန်မာစာ x မြန်မာစာ $P(y=1|x) \cup P(y=2|x)$

5. မြန်မာစာ

- $P(C_1|x) = \frac{P(x|C_1)P(C_1)}{P(x|C_1)P(C_1) + P(x|C_2)P(C_2)} = \frac{1}{1 + \frac{P(x|C_2)P(C_2)}{P(x|C_1)P(C_1)}} = \frac{1}{1 + e^{-z}} = \sigma(z)$
- $\sigma(z) \text{ sigmoid function } [0, 1]$
- $z = wx + b$
 - $w = \ln \frac{P(x|C_1)P(C_1)}{P(x|C_2)P(C_2)}$
 - $b = -\frac{1}{2}(\ln \frac{P(x|C_2)P(C_2)}{P(x|C_1)P(C_1)})^2 + \ln \frac{N_1}{N_2}$
 - $w = \ln \frac{1}{1 + e^{-b}}$
 - $\sigma(w) = \frac{1}{1 + e^{-w}}$

Logistic Regression

Loss function

$$f_{w,b}(x) = P_{w,b}(C_1|x)$$

- မြန်မာစာ

မြန်မာစာ	x^1	x^2	x^3	\dots
C_1	C_1	C_1	C_2	\dots
C_2	$\hat{y} = 1$	$\hat{y} = 1$	$\hat{y} = 0$	\dots

- $f(w, b) = f_{w,b}(x^1)f_{w,b}(x^2)(1-f_{w,b}(x^3))\dots f_{w,b}(x^N)$
- $f_{w,b}(x) \propto C_1 \exp(w^T x) + C_2 \exp(-w^T x)$

$$\ln f_{w,b}(x) = P_{w,b}(C_1|x)$$

$$\ln f_{w,b}(x) = \ln C_1 + \ln \exp(w^T x) + \ln C_2 + \ln \exp(-w^T x)$$

$$\ln P_{w,b}(C_2|x) = 1 - \ln P_{w,b}(C_1|x) = 1 - \ln f_{w,b}(x)$$

- w, b \hat{w}, \hat{b} $\arg \max L(w, b) = \arg \min (-\ln L(w, b))$
- $-\ln L(w, b) = \ln \exp(w^T x) + \ln \exp(-w^T x) + \ln(1 - \exp(w^T x)) = \sum_n [\hat{y}_n \ln \exp(w^T x) + (1 - \hat{y}_n) \ln(1 - \exp(w^T x))]$

- $C(f(x^n), \hat{y}^n) = \sum_n -[\hat{y}^n \ln f_{w,b}(x^n) + (1 - \hat{y}^n) \ln(1 - f_{w,b}(x^n))]$

Gradient Descent

Cross entropy $C(f(x^n), \hat{y}^n) = \sum_n -[\hat{y}^n \ln f_{w,b}(x^n) + (1 - \hat{y}^n) \ln(1 - f_{w,b}(x^n))]$

- $\frac{\partial}{\partial w_i} \frac{\partial \ln(1 - \sigma(z))}{\partial w_i} = \frac{\partial \ln(1 - \sigma(z))}{\partial z} \frac{\partial z}{\partial w_i} = -\sigma(z)x_i$
 - $\frac{\partial}{\partial z} \frac{\partial \ln(1 - \sigma(z))}{\partial z} = -\frac{1}{1 - \sigma(z)}\sigma'(z)$
 - $\frac{\partial}{\partial z} \frac{\partial z}{\partial w_i} = x_i$
- $\frac{\partial}{\partial b} \frac{\partial \ln(1 - \sigma(z))}{\partial b} = \frac{\partial \ln(1 - \sigma(z))}{\partial z} \frac{\partial z}{\partial b} = -(1 - \sigma(z))x_i$
- $\frac{\partial}{\partial w_i} \frac{\partial \ln(-\ln(w, b))}{\partial w_i} = \sum_n -(\hat{y}^n - f_{w,b}(x^n))x_i^n$
 - $\frac{\partial}{\partial w_i} \frac{\partial \ln(-\ln(w, b))}{\partial w_i} = \sum_n -(\hat{y}^n - f_{w,b}(x^n))x_i^n$

Square Error

- **Loss function** $L(f) = \frac{1}{2} \sum_n (f_{w,b}(x^n) - \hat{y}^n)^2$
- $\frac{\partial}{\partial w_i} \frac{\partial L(f)}{\partial w_i} = \text{normalize } 2(f_{w,b}(x) - \hat{y})f_{w,b}(x)(1 - f_{w,b}(x))x_i$
 - $\hat{y} = 0 \Rightarrow f_{w,b}(x) = 1 \Rightarrow 2(f_{w,b}(x) - \hat{y})f_{w,b}(x)(1 - f_{w,b}(x))x_i = 0$
 - $\hat{y} = 1 \Rightarrow f_{w,b}(x) = 0 \Rightarrow 2(f_{w,b}(x) - \hat{y})f_{w,b}(x)(1 - f_{w,b}(x))x_i = 0$

Multi-class Classification

weight · bias

- $C_1: w^1, b_1 \quad z_1 = w^1 \cdot x + b_1$
 - $C_2: w^2, b_2 \quad z_2 = w^2 \cdot x + b_2$
 - $C_3: w^3, b_3 \quad z_3 = w^3 \cdot x + b_3$
- softmax 亂数
 - $f(z) = \frac{e^z}{\sum_{i=1}^n e^{z_i}}$
 - $f(z) \in [0, 1]^n$
- Cross entropy 亂数
 - $\frac{1}{n} \sum_{i=1}^n -\hat{y}_i \ln y_i$

Limitation of Logistic Regression

- 亂数
- $\sigma(z) = 0.5 \Rightarrow w \cdot x + b = 0$

Feature transformation

Discriminative and Generative

Discriminative

• ပုံစံမျက်နှာတွင် x အတွက်

- $P(y|x)$

- ပုံစံမျက်နှာတွင် y အတွက်

Generative ပုံစံမျက်နှာ

• ပုံစံမျက်နှာတွင် $P(x|y)$

- ပုံစံမျက်နှာတွင် x အတွက် $w \cdot b$

- ပုံစံမျက်နှာတွင် y အတွက်

Deep Learning

Three steps for Deep Learning

1. Define a set of function

- ပုံစံမျက်နှာတွင်

- ပုံစံ

- ပုံစံမျက်နှာတွင် σ အတွက်

- ပုံစံမျက်နှာတွင် ReLU・Sigmoid・Tanh အဲ

- ပုံစံမျက်နှာတွင် Adam အတွက်

- ပုံစံမျက်နှာတွင် SGD အတွက်

ပုံစံ

- ပုံစံမျက်နှာတွင် Sigmoid အတွက် ReLU・ReLU အတွက်

Sigmoid ပုံစံ

- $\sigma(x) = (0, 1)$

- ပုံစံမျက်နှာတွင် 0 အတွက်

- ပုံစံမျက်နှာတွင်

ReLU ပုံစံ

- $\text{ReLU}(x) = [0, \infty)$

- ပုံစံမျက်နှာတွင် sigmoid ပုံစံမျက်နှာတွင်

- ပုံစံမျက်နှာတွင်

□ ReLU ပုံစံ

- ReLU ပုံစံမျက်နှာတွင် 0 အတွက်

2. Goodness of function

- ပုံစံမျက်နှာတွင်

- ပုံစံမျက်နှာတွင် σ အတွက်

- ပုံစံမျက်နှာတွင် k အတွက်

- ပုံစံမျက်နှာတွင် ReLU အတွက်

- ပုံစံမျက်နှာတွင် SGD အတွက်

3. Pick the best function

- $\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
- MSE ကို k ပုံမှန်လေ့လာရန် အသင့်ဆုံး ဖြစ်တယ်
- MSE ကို k ပုံမှန်လေ့လာရန် အသင့်ဆုံး ဖြစ်တယ်
- MSE ကို k ပုံမှန်လေ့လာရန် အသင့်ဆုံး ဖြစ်တယ်

Gradient Descent

- Learning Rate α
 - α ကို အမြင်လေ့လာ
 - α ကို အမြင်လေ့လာ

ဗိုလ်ချုပ် Taylor Series အတွက်

Adagrad

ဗိုလ်ချုပ်

Stochastic Gradient Descent

ဗိုလ်ချုပ် $L = \sum_n (\hat{y}^n - (b + \sum w_i x_i^n))^2$

- Gradient Descent $\theta^i = \theta^{i-1} - \eta \nabla L(\theta^{i-1})$
- Stochastic Gradient Descent $\theta^i = \theta^{i-1} - \eta \nabla L(\theta^{i-1})$ example $\theta^i = \theta^{i-1} - \eta \nabla L(\theta^{i-1})$
 - $\nabla L(\theta^{i-1})$ ကို ဘယ်လေ့လာမှု

Feature Scaling

ဗိုလ်ချုပ်

- $x_i \rightarrow \frac{x_i - m_i}{\sigma_i}$
- $x_i \rightarrow \frac{x_i - m_i}{\sigma_i}$

ဗိုလ်ချုပ်

- $x_i \rightarrow \frac{x_i - m_i}{\sigma_i}$
 - m_i \$ σ_i \$
 - 0 1

Backpropagation

- Forward pass \Rightarrow Backword pass
 - $\text{Forward pass} \Rightarrow \text{Backword pass}$
 - $\text{Forward pass} \Rightarrow \text{Backword pass}$
 - $\text{Forward pass} \Rightarrow \text{Backword pass}$



Forward pass

Backword pass

BERT

BERT

Unsupervised Learning

Self-supervised Learning

Word Embedding

Word Embedding

Word

1. One-hot encoding

- 一个词的独热编码***(one-hot encoding)***是将该词表示为一个向量，其中只有一个元素为1，其余均为0。
- 向量维度通常在50-300之间。

2. Word2Vec

- 使用滑动窗口技术来捕获上下文信息。
- 使用向量表示词义。

Prediction-based

Prediction-based

预测基于

预测

1. 通过独热编码进行one-hot encoding

- "狗" "猫" "鸟" "狗" "鸟" "狗" "鸟" "狗" "鸟"

2. 通过矩阵乘法进行

- " $\begin{pmatrix} \text{狗} \\ \text{猫} \end{pmatrix} \times \begin{pmatrix} \text{狗} & \text{猫} & \text{鸟} \end{pmatrix} \rightarrow \begin{pmatrix} \text{狗} \\ \text{猫} \\ \text{鸟} \end{pmatrix}$ "
- " $\begin{pmatrix} \text{狗} \\ \text{猫} \end{pmatrix} \times \begin{pmatrix} \text{狗} & \text{猫} & \text{鸟} \end{pmatrix} \rightarrow \begin{pmatrix} \text{狗} \\ \text{猫} \\ \text{鸟} \end{pmatrix}$ "
- "狗" "猫" "鸟" "狗" "猫" "鸟" "狗" "猫" "鸟"
- "狗" "猫" "鸟" "狗" "猫" "鸟" "狗" "猫" "鸟"

4. 通过神经网络进行

- 神经网络模型进行预测

- ပုဂ္ဂန်များ

5. အောက်တို့တော်းခြားထားရေးနည်းလမ်း

- ပုဂ္ဂန်များ
- ပုဂ္ဂန်များ

အောက်တို့တော်းခြားထားရေးနည်းလမ်း

1. အောက်တို့တော်းခြားခြင်း

- ပုဂ္ဂန်များ
- ပုဂ္ဂန်များ

2. အောက်တို့တော်းခြားခြင်း

- ပုဂ္ဂန်များ
- ပုဂ္ဂန်များ

Seq2Seq

Sequence to Sequence Modelအောက်တို့တော်းခြားခြင်း

အောက်

Seq2Seqအောက်တို့တော်းခြားခြင်း

1. အောက်Encoderအောက်တို့တော်းခြားခြင်း

2. အောက်Decoderအောက်တို့တော်းခြားခြင်း

AT Auto-regressive VS **NAT** Non-Autoregressive Transformerအောက်

အောက်**AR/AT**

- အောက်တို့တော်းခြားခြင်း
- အောက်တို့တော်းခြင်း
- အောက်တို့တော်းခြင်း
- အောက်GPTအောက်Transformerအောက်

အောက်**NAT**

- အောက်တို့တော်းခြင်း
- အောက်တို့တော်းခြင်း
- အောက်တို့တော်းခြင်း
- အောက်တို့တော်းခြင်း