Machine-Learning-Model-Analysis Documentation

Release 0.0.1

476900563@qq.com

Mar 04, 2019

Contents

CHAPTER 1

mathematical notation

- x
- $\bar{x}$
- $\mathbf{x}$
- $\mathbf{x}^T$
- $\mathbf{X}$,,,
- n
- m
- $\mathbf{x}_i$ i
- $\mathbf{X}_i$ i
- $\mathbf{X}^j$ j
- $\mathbf{X}_i^j$ ij
- ϵ
- ;
- whl

CHAPTER 2

linear Model

2.1 linear regression

2.1.1

2.1.2 Predicted function

$$f(\mathbf{x}) = \begin{bmatrix} \theta_0 \\ \theta_1 \\ .. \\ \theta_n \end{bmatrix}^T \begin{bmatrix} \mathbf{x}_0 \\ \mathbf{x}_1 \\ .. \\ \mathbf{x}_n \end{bmatrix} = \theta^T \mathbf{x}$$

2.1.3 Loss function

$$L(y, \bar{y}) = (y - \bar{y})^2$$

2.1.4 Object function

$$O(\mathbf{y}, \mathbf{X}) = \frac{1}{2} \sum_{i=1}^{m} L(\mathbf{y}_i, f(\mathbf{X}_i)) = (\mathbf{y} - \mathbf{X}\theta)^T (\mathbf{y} - \mathbf{X}\theta)$$

2.1.5 Optimizing

$$\theta = \arg\min_{\theta} \ \frac{1}{2} O(\mathbf{y}, \mathbf{X})$$

Normal equations

$$\begin{aligned}\nabla_\theta O(\mathbf{y},\mathbf{X};\theta) &= \nabla_\theta \ \frac{1}{2}(\mathbf{y}-\mathbf{X}\theta)^T(\mathbf{y}-\mathbf{X}\theta)\\ &= \frac{1}{2}\nabla_\theta \ (\theta^T\mathbf{X}^T\mathbf{X}\theta-\theta^T\mathbf{X}\mathbf{y}-\mathbf{y}^T\mathbf{X}\theta+\mathbf{y}^T\mathbf{y})\\ &= \frac{1}{2}\nabla_\theta \ \mathrm{Tr}(\theta^T\mathbf{X}^T\mathbf{X}\theta-\theta^T\mathbf{X}\mathbf{y}-\mathbf{y}^T\mathbf{X}\theta+\mathbf{y}^T\mathbf{y})\\ &= \frac{1}{2}\nabla_\theta \ (\mathrm{Tr}(\theta^T\mathbf{X}^T\mathbf{X}\theta)-2\mathrm{Tr}(\mathbf{y}^T\mathbf{X}\theta))\\ &= \frac{1}{2}(\mathbf{X}^T\mathbf{X}\theta+\mathbf{X}^T\mathbf{X}\theta-2\mathbf{X}^T\mathbf{y})\\ &= \mathbf{X}^T\mathbf{X}\theta-\mathbf{X}^T\mathbf{y}\end{aligned}$$

$\nabla_\theta O(\mathbf{y},\mathbf{X};\theta)=0$,

$$\begin{aligned}\mathbf{X}^T\mathbf{X}\theta &= \mathbf{X}^T\mathbf{y}\\ \theta &= \begin{cases}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} & \mathrm{Tr}\mathbf{X}=\mathrm{Tr}\mathbf{X}^T\\ (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} & \text{pass}\end{cases}\end{aligned}$$

Warning: ,

Note:

Gradient-based

$$\begin{aligned}\nabla_\theta O(\mathbf{y},\mathbf{X};\theta) &= \nabla_\theta \ \frac{1}{2}\sum_{i=1}^{m} L(\mathbf{y}_i, f(\mathbf{X}_i))\\ &= \nabla_\theta \ \frac{1}{2}\sum_{i=1}^{m}(\mathbf{y}_i-\theta^T\mathbf{X}_i)^2\\ &= \nabla_\theta \ \frac{1}{2}(\mathbf{y}_i-\theta^T\mathbf{X}_i)^2\\ &= \frac{1}{2}*2(\mathbf{y}_i-\theta^T\mathbf{X}_i)\frac{\partial(-\theta^T\mathbf{X})}{\mathrm{d}\theta}\\ &= -(\mathbf{y}_i-\theta^T\mathbf{X}_i)\mathbf{X}_i\end{aligned}$$

,:

$$\theta \leftarrow \theta-\alpha(-(\mathbf{y}_i-f(\mathbf{X}_i))\mathbf{X}_i)$$

2.2 linear regression why

2.2.1 ?

: . $P(y|x;\theta) \sim \mathcal{N}(\mu, \sigma^2)$.:

$$\begin{aligned} P(y;\mu) =& \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(y-\mu)^2) \\ =& \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2). \exp(\mu y - \frac{1}{2}\mu^2) \end{aligned}$$

: $\eta = \mu$.,:

$$\begin{aligned} f(x) &= \mathbf{E}[y|x;\theta] \\ &= \mu \end{aligned}$$

$\eta = \theta^T \mathbf{x}$. ,:

$$f(x) = \theta^T \mathbf{x}$$

2.2.2 ?

: $\epsilon \sim \mathcal{N}(0, \sigma^2)$. :

$$y = \theta^T \mathbf{x} + \epsilon$$

:

$$P(y : x; \theta) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{(y-\theta^T \mathbf{x})^2}{2\sigma^2})$$

:

$$\begin{aligned} L(\theta) =& \prod_{i=1}^{m} P(\mathbf{y}_i|\mathbf{X}_i;\theta) \\ \ell =& \log \prod_{i=1}^{m} P(\mathbf{y}_i|\mathbf{X}_i;\theta) \\ =& \log \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\mathbf{y}_i - \theta^T \mathbf{X}_i)^2}{2\sigma^2}\right) \\ =& \sum_{i=1}^{m} \log - \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(\mathbf{y}_i - \theta^T \mathbf{X}_i)^2}{2\sigma^2}) \\ =& m \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2} \cdot \frac{1}{2} \sum_{i=1}^{m} (\mathbf{y}_i - \theta \mathbf{X}_i)^2 \\ \approx& -\frac{1}{2} \sum_{i=1}^{m} (\mathbf{y}_i - \theta \mathbf{X}_i)^2 \end{aligned}$$

.

2.3 linear regression expand

2.3.1 Lasso

2.3.2 Ridge

2.3.3 LWR(locally weighted linear regression)

2.4 logistic regression

2.4.1

2.4.2 Predicted function

$$f(x) = g(\theta^T \mathbf{x}) = \frac{1}{1 + \exp -\theta^T \mathbf{x}}$$

2.4.3 Loss function

$$L(y, \bar{y}) = -(y \log f(x) + (1 - y) \log (1 - f(x)))$$

Note: ,log(2+exp)

Note:

2.4.4 Objection function

$$O(\mathbf{y}, \mathbf{X}) = \sum_{i=1}^{m} L(\mathbf{y}_i, f(\mathbf{X}_i) = -\sum_{i=1}^{m} \mathbf{y}_i \log f(\mathbf{X}_i) + (1 - \mathbf{y}_i) \log (1 - f(\mathbf{X}_i))$$

2.4.5 Optimizing

$$\theta = \arg\min_{\theta} \ O(\mathbf{y}, \mathbf{X})$$

Gradient-based

$$\begin{aligned}
\nabla_\theta O(\mathbf{y}, \mathbf{X}; \theta) &= \nabla_\theta \sum_{i=1}^{m} L(\mathbf{y}_i, f(\mathbf{X}_i)) \\
&= \nabla_\theta - \sum_{i=1}^{m} \mathbf{y}_i \log f(\mathbf{X}_i) + (1 - \mathbf{y}_i) \log (1 - f(\mathbf{X}_i)) \\
&= \nabla_\theta - (\mathbf{y}_i \log f(\mathbf{X}_i) + (1 - \mathbf{y}_i) \log (1 - f(\mathbf{X}_i)) \\
&= -\frac{\partial \mathbf{y}_i \log f(\mathbf{X}_i) + (1 - \mathbf{y}_i) \log (1 - f(\mathbf{X}_i))}{\mathrm{d}(\theta^T \mathbf{X}_i)} \frac{\partial \theta^T \mathbf{X}_i}{\mathrm{d}x} \\
&= -(\mathbf{y}_i \frac{1}{f(\mathbf{X}_i)} + (1 - y) \frac{1}{1 - f(\mathbf{X}_i)}) g(\theta^T \mathbf{X}_i)(1 - g(\theta^T \mathbf{X}_i))\mathbf{X}_i \\
&= -(\mathbf{y}_i - f(\mathbf{X}_i))\mathbf{X}_i
\end{aligned}$$

,:

$$\theta \leftarrow \theta - \alpha(-(\mathbf{y}_i - f(\mathbf{X}_i))\mathbf{X}_i)$$

2.5 logistic regression why

2.5.1 ?

. $P(y|x;\theta) \sim \text{Bernoulli}(\phi)$.:

$$\begin{aligned}
P(y;\phi) =& \phi^y (1 - \phi)^{1-y} \\
=& \exp\left(y \log \phi + (1 - y) \log (1 - \phi)\right) \\
=& \exp\left((\log (\frac{\phi}{1 - \phi}))y + \log (1 - \phi)\right)
\end{aligned}$$

: $\eta = \log \frac{\phi}{1-\phi}$. ,:

$$\begin{aligned}
f(x) &= \mathbf{E}[y|x;\theta] \\
&= \phi
\end{aligned}$$

$\phi = \theta^T \mathbf{x}$. ,:

$$f(x) = \frac{1}{1 + \exp -\theta^T \mathbf{x}}$$

2.5.2 ?

(odd):

$$\log \frac{y}{1 - y} = \theta^T \mathbf{x}$$

:

$$\begin{aligned}
P(y = 1|x;\theta) &= f(\mathbf{x}) \\
P(y = 0|x;\theta) &= 1 - f(\mathbf{x})
\end{aligned}$$

:

$$P(y|x;\theta) = f(\mathbf{x})^y(1-f(\mathbf{x}))^{1-y}$$

:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{m} P(\mathbf{y}_i|\mathbf{X}_i;\theta) \\ \ell &= \log \prod_{i=1}^{m} P(\mathbf{y}_i|\mathbf{X}_i;\theta) \\ &= \log \prod_{i=1}^{m} f(\mathbf{X}_i)^{\mathbf{y}_i}(1-f(\mathbf{X}_i))^{1-\mathbf{y}_i} \\ &= \sum_{i=1}^{m} \mathbf{y}_i \log f(\mathbf{X}_i) + (1-\mathbf{y}_i)\log\left(1-f(\mathbf{X}_i)\right) \end{aligned}$$

.

2.6 logistic regression expand

2.7 softmax regression

2.7.1

2.7.2 Predicted function

$$f(\mathbf{x}, k; \mathbf{k}) = \frac{\exp^{\mathbf{x}\boldsymbol{\Theta}_k}}{\sum\limits_{k=1}^{\mathbf{k}} \exp^{\mathbf{x}\boldsymbol{\Theta}_k}}$$

Note: ,

2.7.3 Loss function

$$L(\mathbf{y}, \bar{\mathbf{y}}) = -\prod_{k=1}^{\mathbf{k}} \mathbf{y}_k \log \bar{\mathbf{y}}_k$$

2.7.4 Object function

$$\begin{aligned} O(\mathbf{X}, \mathbf{Y}; \boldsymbol{\Theta}) &= \prod_{i=1}^{m} L(\mathbf{Y}_i, \bar{\mathbf{Y}}_i) \\ &= -\prod_{i=1}^{m}\prod_{k=1}^{\mathbf{k}} \mathbf{Y}_i^k \log \bar{\mathbf{Y}}_i^k \end{aligned}$$

2.7.5 Optimizing

Gradient-based

$$\nabla_\Theta O(\mathbf{y},\mathbf{X};\mathbf{\Theta}) = -\nabla_\Theta \prod_{i=1}^{m}\prod_{k=1}^{\mathbf{k}} \mathbf{Y}_i^k \log \bar{\mathbf{Y}}_i^k \tag{2}$$

$$= -\nabla_\Theta \prod_{i=1}^{m}\prod_{k=1}^{\mathbf{k}} \mathbf{Y}_i^k \log f(\mathbf{X}_i,k;\mathbf{k}) \tag{3}$$

$$= -\nabla_\Theta \prod_{k=1}^{\mathbf{k}} \mathbf{Y}_i^k \log f(\mathbf{X}_i,k;\mathbf{k}) \tag{4}$$

$$= -\nabla_{\Theta_j} \prod_{k=1}^{\mathbf{k}} \mathbf{Y}_i^k \log f(\mathbf{X}_i,k;\mathbf{k}) \tag{5}$$

$$= -\mathbf{Y}_i^k \frac{\partial \log f(\mathbf{X}_i,k;\mathbf{k})}{\mathrm{d}f(\mathbf{X}_i,k;\mathbf{k})} \frac{\partial f(\mathbf{X}_i,k;\mathbf{k})}{\mathrm{d}(\Theta_j\mathbf{X}_i)} \frac{\partial(\Theta_j\mathbf{X}_i)}{\mathrm{d}\Theta_j} \tag{6}$$

$$= -\mathbf{Y}_i^k \frac{1}{f(\mathbf{X}_i,k;\mathbf{k})} \frac{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})(\exp^{\Theta_k\mathbf{X}_i})' - (\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})'(\exp^{\Theta_k\mathbf{X}_i})}{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})^2}\mathbf{X}_i \tag{7}$$

$$= -\begin{cases} \mathbf{Y}_i^k \frac{1}{f(\mathbf{X}_i,j;\mathbf{k})} \frac{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})(\exp^{\Theta_j\mathbf{X}_i}) - (\exp^{\Theta_j\mathbf{X}_i})(\exp^{\Theta_j\mathbf{X}_i})}{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})^2}\mathbf{X}_i & j == k \\ \mathbf{Y}_i^k \frac{1}{f(\mathbf{X}_i,k;\mathbf{k})} \frac{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})\times 0 - ((\exp^{\Theta_k\mathbf{X}_i})(\exp^{\Theta_k\mathbf{X}_i})}{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})^2}\mathbf{X}_i & j \neq k \end{cases} \tag{8}$$

$$= -\begin{cases} \mathbf{Y}_i^k \frac{1}{\exp^{\Theta_j\mathbf{X}_i}} \frac{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})(\exp^{\Theta_j\mathbf{X}_i}) - (\exp^{\Theta_j\mathbf{X}_i})(\exp^{\Theta_j\mathbf{X}_i})}{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})}\mathbf{X}_i & j == k \\ \mathbf{Y}_i^k \frac{1}{\exp^{\Theta_k\mathbf{X}_i}} \frac{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})\times 0 - ((\exp^{\Theta_k\mathbf{X}_i})(\exp^{\Theta_k\mathbf{X}_i})}{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})}\mathbf{X}_i & j \neq k \end{cases} \tag{9}$$

$$= -\begin{cases} \mathbf{Y}_i^k \frac{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i}) - (\exp^{\Theta_j\mathbf{X}_i})}{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})}\mathbf{X}_i & j == k \\ \mathbf{Y}_i^k \frac{-(\exp^{\Theta_k\mathbf{X}_i})}{(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})}\mathbf{X}_i & j \neq k \end{cases} \tag{10}$$

$$= -\begin{cases} \mathbf{Y}_i^k(1 - f(\mathbf{x},k;\mathbf{k}))\mathbf{X}_i & j == k \\ \mathbf{Y}_i^k - f(\mathbf{x},k;\mathbf{k})\mathbf{X}_i & j \neq k \end{cases} \tag{11}$$

Note: $(\exp^{\Theta_k\mathbf{X}_i})'$ $(\sum_{k=1}^{m}\exp^{\Theta_k\mathbf{X}_i})'$ $\Theta_k\mathbf{X}_i$.

2.8 softmax regression why

2.8.1 ?

. $P(\mathbf{y}|x;\phi,k) \sim \text{category}(\phi_1,\phi_2,...,\phi_i,...,\phi_k)$.:

$$\begin{aligned}
P(\mathbf{y};\theta,k) &= {\phi_1}^{\mathbf{y}_1}{\phi_2}^{\mathbf{y}_2}...{\phi_i}^{\mathbf{y}_i}...{\phi_k}^{1-\sum_{i=1}^{k-1}\mathbf{y}} \\
&= \exp^{\mathbf{y}_1\log\phi_1+\mathbf{y}_2\log\phi_2+...+\mathbf{y}_i\log\phi_i+...+(1-\sum_{i=1}^{k-1}\mathbf{y})\log\phi_k} \\
&= \exp^{\mathbf{y}_1\log\frac{\phi_1}{\phi_k}+\mathbf{y}_2\log\frac{\phi_2}{\phi_k}+...+\mathbf{y}_{k-1}\log\frac{\phi_{k-1}}{\phi_k}+\log\phi_k} \\
&= \exp\begin{bmatrix}\log\frac{\phi_1}{\phi_k}\\ \log\frac{\phi_2}{\phi_k}\\ ...\\ \log\frac{\phi_{k-1}}{\phi_k}\\ \log\frac{\phi_k}{\phi_k}\end{bmatrix}^T \mathbf{y} - -\log\phi_k
\end{aligned}$$

: $\eta = \begin{bmatrix}\log\frac{\phi_1}{\phi_k}\\ \log\frac{\phi_2}{\phi_k}\\ ...\\ \log\frac{\phi_{k-1}}{\phi_k}\\ \log\frac{\phi_k}{\phi_k}\end{bmatrix}$. :

$$\begin{aligned}
\eta_i &= \log\frac{\phi_i}{\phi_k} \\
\phi_i &= \phi_k\exp^{\eta_i} \\
1 &= \sum_{i=1}^{k}\phi_i = \phi_k\sum_{i=1}^{k}\exp^{\eta_i} \\
\phi_k &= \frac{1}{\sum_{i=1}^{k}\exp^{\eta_i}}
\end{aligned}$$

$\phi_k = \frac{1}{\sum_{i=1}^{k}\exp^{\eta_i}}$ $\phi_i = \phi_k\exp^{\eta_i}$,:

$$\phi_i = \frac{\exp^{\eta_i}}{\sum_{i=1}^{k}\exp^{\eta_i}}$$

:

$$\begin{aligned}
f(x) &= \mathbf{E}[\mathbf{y}|x;\phi] \\
&= \begin{bmatrix}\phi_1\\ \phi_2\\ ...\\ \phi_k\end{bmatrix}
\end{aligned}$$

$\phi_i = \theta_i^T \mathbf{x}$. ,:

$$f(x) = \begin{bmatrix} \frac{\exp^{\theta_1^T \mathbf{x}}}{\sum_{i=1}^{k} \exp^{\theta_i^T \mathbf{x}}} \\ \frac{\exp^{\theta_2^T \mathbf{x}}}{\sum_{i=1}^{k} \exp^{\theta_i^T \mathbf{x}}} \\ \cdots \\ \frac{\exp^{\theta_k^T \mathbf{x}}}{\sum_{i=1}^{k} \exp^{\theta_i^T \mathbf{x}}} \end{bmatrix}$$

Note: $1 \leq i \leq k$

CHAPTER 3

Neruo layers

3.1 convolution layer

3.1.1

3.1.2

$$O(n) = \sum_{i=n}^{w+n} I(i)K(i)$$

$$\frac{\partial O(n)}{\mathrm{d}K} = I[n : w + n]$$

$$\frac{\partial(\sum_{n=0}^{N} O(n))}{\mathrm{d}K} = \sum_{n=0}^{o} I[n : w + n]$$

$$O(m, n) = \sum_{i=m}^{w+m} \sum_{j=n}^{h+n} I(i, j)K(i, j)$$

$$\frac{\partial O(m, n)}{\mathrm{d}K} = I[m : w + m; n : h + n]$$

$$\frac{\partial(\sum_{m=0}^{M}\sum_{n=0}^{N}O(m,n))}{\mathrm{d}K}=\sum_{m=0}^{M}\sum_{n=0}^{N}I[m:w+m;n:h+n]$$

$$O(m,n,v)=\sum_{i=m}^{w+m}\sum_{j=n}^{h+n}\sum_{k=v}^{l+v}I(i,j,k)K(i,j,k)$$

$$\frac{\partial O(m,n,v)}{\mathrm{d}K}=I[m:w+m;n:h+n;v:l+v]$$

$$\frac{\partial(\sum_{m=0}^{M}\sum_{n=0}^{N}\sum_{v=0}^{V}O(m,n,v))}{\mathrm{d}K}=\sum_{m=0}^{M}\sum_{n=0}^{N}\sum_{v=0}^{V}I[m:w+m;n:h+n;v:l+v]$$

Warning:

3.1.3

3.2 Full connection layer

3.2.1

3.2.2

()

$$O(n)=\sum_{i=1}^{x}I(i)K(i)_n$$

$$\frac{\partial O(n)}{\mathrm{d}K_n}=I$$

()

$$O(n) = \sum_{i=1}^{x}\sum_{j=1}^{y} I(i,j)K(i,j)_n$$

$$\frac{\partial O(n)}{\mathrm{d}K_n} = I$$

()

$$O(n) = \sum_{i=1}^{x}\sum_{j=1}^{y}\sum_{k=1}^{z} I(i,j,k)K(i,j,k)_n$$

$$\frac{\partial O(n)}{\mathrm{d}K_n} = I$$

3.2.3

3.3 pooling layer

3.3.1

3.3.2

$$O(m) = \sum_{i=m}^{w+m} I(i)K(i)$$

$$K(i) = \begin{cases} 1 & if \;\; I(i) \text{is maximum} \\ 0 & else \;\; \text{other} \end{cases}$$

$$O(m,n) = \sum_{i=m}^{w+m}\sum_{j=n}^{h+n} I(i,j)K(i,j)$$

$$K(i,j) = \begin{cases} 1 & if \;\; I(i,j) \text{is maximum} \\ 0 & else \;\; \text{other} \end{cases}$$

$$O(m,n) = \sum_{i=m}^{w+m} \sum_{j=n}^{h+n} \sum_{k=z}^{l+z} I(i,j,k)K(i,j,k)$$

$$K(i,j,k) = \begin{cases} 1 & if \;\; I(i,j,k) \text{is maximum} \\ 0 & else \;\; \text{other} \end{cases}$$

$$O(m) = \sum_{i=m}^{w+m} I(i)K(i)$$

$$k(i) = \frac{1}{w}$$

$$O(m,n) = \sum_{i=m}^{w+m} \sum_{j=n}^{h+n} I(i,j)K(i,j)$$

$$k(i,j) = \frac{1}{w \times h}$$

$$O(m,n) = \sum_{i=m}^{w+m} \sum_{j=n}^{h+n} \sum_{k=z}^{l+z} I(i,j,k)K(i,j,k)$$

$$k(i,j,k) = \frac{1}{w \times h \times l}$$

3.3.3

-
-
- ()

CHAPTER 4

neural network model

4.1 LeNet-5

4.1.1

4.1.2

4.2 MLP

4.2.1

4.2.2

CHAPTER 5

Todo

- rst Ok
- Machine-Learning-Model-Analysis ok
- ok
- gradle ok
- ,.
-
- linear regression expand (lassoridge)
- logistic
- logisticIRLS
-
-
-
- ()
-

5.1 MxNet-Hands-On-Deep-Learning

5.1.1 ()

-
- logistic
- softmax

5.1.2 :GPU()

- MLP ,
- ?
- ?
- LeNet ,
- —-
-

CHAPTER 6

Indices and tables

- genindex
- modindex
- search