Экзамен по машинному обучению ФПМИ[1]

Список тем

1. Метрики качества классификации: accuracy, balanced accuracy, precision, recall, f1-score (/w multiclass extensions), ROC-AUC.

Предположим, что рассматривается задача бинарной классификации (т.е. с классами $\{-1, +1\}$). Определим следующие величины:

- TP (True positive) — количество правильно угаданных меток позитивного класса (обычно класса с меткой $+1$).
- TN (True negative) — количество правильно угаданных меток негативного класса (обычно класса с меткой 0 или -1).
- FP (False positive) — количество объектов негативного класса, которые классификатор назвал позитивными.
- FN (False negative) — количество объектов позитивного класса, которые классификатор назвал негативными.

		True condition	
	Total population	Condition positive	Condition negative
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error
	Predicted condition negative	**False negative**, Type II error	**True negative**

Список метрик:

- **Accuracy** (точность) — количество правильно классифицированных объектов:

$$\frac{1}{n}\sum_{i=1}^{n}[y_i = y_i^p]$$

  Здесь y_i^p — предсказанная метка, а y_i — истинная.
  - **Balanced accuracy** — усреднённая по классам точность

$$\frac{1}{C}\sum_{k=1}^{C}\frac{\sum_i[y_i = y_i^p \text{ and } y_i = k]}{\sum_i[y_i = k]}$$

- **Ē Ţ Ĥ ĺ ą̄ k ā' ę** ĥ ÿ ĥ Ĝ ě ĥ Ū ÿ γ ı 'A Ò ĥ ÿ σ *ìă ãåõùÕõy* ù ÿ Ē Ū Ū ù Ř ù Γ ù ö ĥ ě Ē ě ě Y ř ĥ è T ố ù ÿ ĥ ể ù ở ÿ ố ř ĥ ù ĥ ÿ ĥ ö Y ố ù ÿ Ē Ū Ū ù Ř ù ù Ē ÿ ĥ ö ě Ē ỏ ể Ē ÿ ī ĥ ỏ ù ÿ ù ể ě Y ẽ ù ĥ

$$\frac{TP}{TP + FP}$$

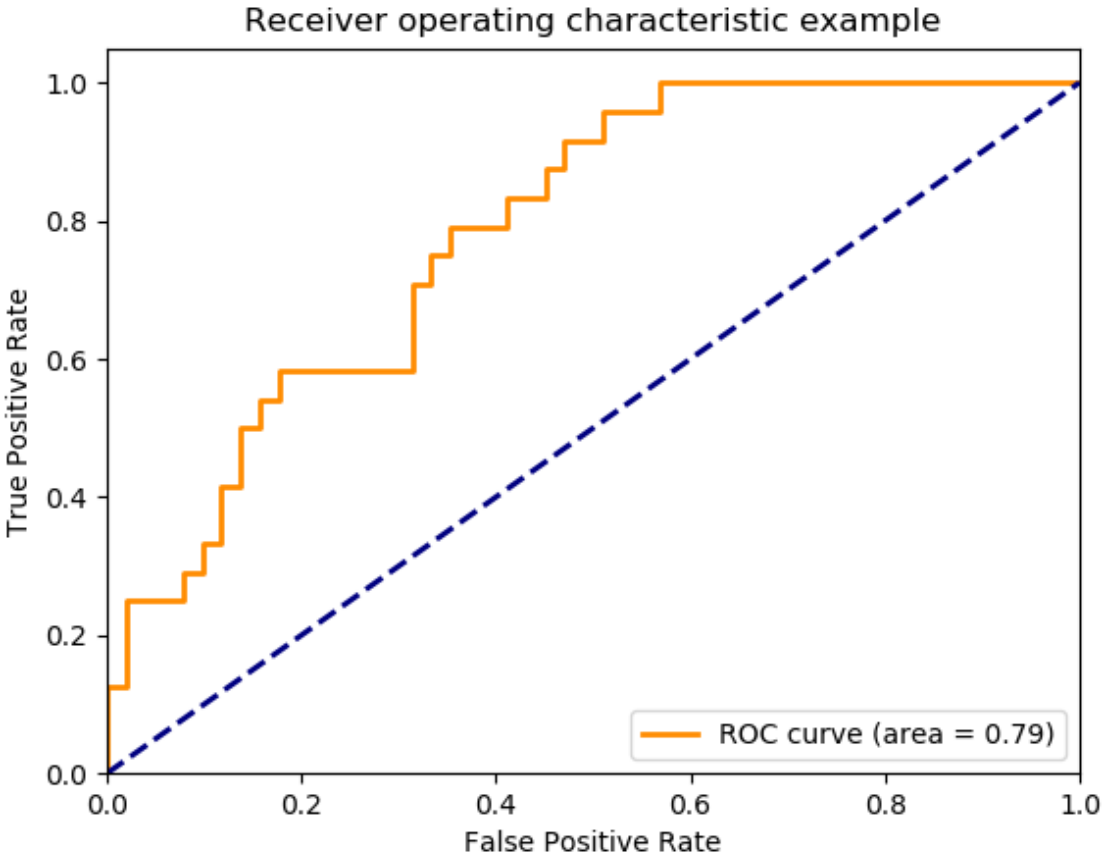
- **Ė X ĺ È È** ī ĥ ÿ ě ĥ ÿ Ē ı 'A Ò ĥ ÿ σ ĥ è T ố ù ÿ ĥ ể ī ĥ ỏ ù ÿ ù ể ě ĥ ẽ ĥ ù ÿ Ē Ū Ū Ē ĥ ù ĥ ÿ ĥ ö Y ř ī ö Ē ể ù ÿ γ ě ĥ ù ÿ Ē Ū Ū ù Ř ù Γ ù ö ĥ ể Ē ÿ Ē ÿ ẽ ĥ ö ù ÿ ẽ ĥ

$$\frac{TP}{TP + FN}$$

- **ô⁵ ą̄ k ĺ " Ţ Ĥ** 'A ẽ Ē ö ẽ ĥ ě ù Θ ố Ū ù ĥ ố Ū ö ố Ò ě ố ố *ḍ è ẽ Ĝ Ĝ Ĝ

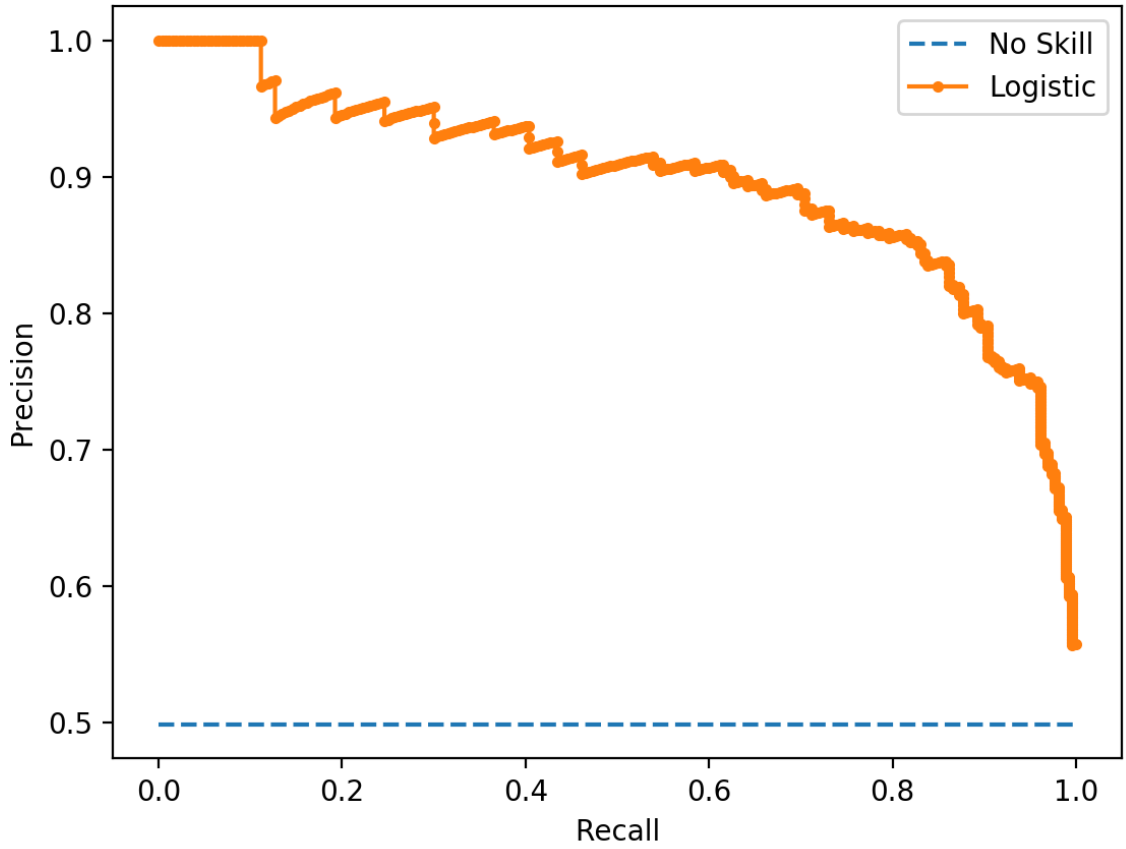
- pěĚŒőěụố 2ń b ”IĖ˙ b ụẽ ốốỹ Ỹẽ YỸy ểốö ẖơỹể ẖỸỹụ ỹẖẽẖ ẖỦẖỹ ẖốỸyụ ề Yyụ ểYề ö Ẽể Y Ỹ ỵ ỹ ŒẼỦể YỦ ỉẖ ẖỹ ẖỨụỹốỵ ỵể YỦ ụ Ỹ ỵ ỹ ŒẼỦể YỦ ẖỹ ö ụ ỦẼốỵ ỵể YỦ ẖề Tốủỹ Y ểYề ẖ ö ủ ụ ỉ ỉẖ ỉẖ ỹẖỨụỹốỵ ỵể YỦ ẖề Tốủỹ ỉẖ ỹ ỹ Œụỹ ẖỦ ốể ủỹ ỉ ö ụ ể Ẽỏỵ ốỨể ẖỸỹụ ểYỦ ốỉẖ Œốẽ ẖỹ ö ụ ỦẼỹốỵ ỵể YỦ ẖề Tốủỹỉẖ

- ‘ öụẽ ốö Ė˙ b ủ ủ ö ụ ể ẖỦH



- **ĒĖ ạ ốỵ X** ’A ẽ ö Ẽ Ř ụ ử ể ẖỸơ ĒỵẖỊ ạ ả ẻ ẹ ẖỸỹ y_i ụ ĖỊ ỀỀẖỸỹ x_i ỉ ỉ
  - " ỹ ö ẖ ụ ỹỸơ Ẽể Ẽ ỵẖẽ ụ Œể ẖ Ė˙ b ủ ủ ö ụ ể ẖỦỉỉỸ ỉẖẽ ẖ Σ γμ ỉ ốö ốẽ ốΣ ốể ụ σẽ ö Ẽể ụ Γ Yỉ ỉ
  - " ể ẖỦỸỹể Ẽ˙H
    - LỦốẽỎẼ ụỎỎỹ ụ ở ỹẖ ủ ụ $(0,1)$ ể ỹẖ ủ ч $(1,r)$, ẽỎố r ’A Ỏẖỹ σẖề Tốủỹẖể ỉẖ ở ụỹ ụể ể ẖẽẖ ủỵ ẼỸỸẼỉ ỉ
    - ĒĖ ủ ủ ö ụ ể Ẽ σỸỵ ч ŒẼỦể ẖỦ ẽ ẖỎốỵ ụ ỵốỨụỹ ể Ẽẖỹ ụ xỉ ỉ
    - ₌ ốẽ ể YΠố ủ ö ụ ể Ẽ σỹốẽ ỵ ч ŒΠố ẽẖỎốỵ ỵ
    - x ỎốẼỵ ỵể Ẽ σẽẖỎốỵ ỵ ỉ ö ẖ ẖỎụỹ Œ ố ố ở ỹẖ ủ ч $(1,1.)$
  - **ĒĖ ố 2ń b** ’A ỉỵẖΣ ẼỎỵ ỉẖỎ ĒĖ ủ ủ ö ụ ể ẖỦỉỉ

- ‘ öụẽ ốö ẽẺẻửöụểẻửH



⁶ ỉỉЛъốÿöụửụ ửẽẽỐốỸÿểẽẽ öốẽöốỸỸụụ'Ĥ ̀ g t ỉỉÎ ̀ 2 t ỉỉẺ⁶ ỉỉỎöчẽụốỉỉ

pỎốỸγ $\hat{Y}$ ụ $\hat{y}$ 'A ïöốỎỸửẽỏẽẽểẽYố ỏểẽẽỐốểụσÿẽöẽốÿẽỉỉY ụ y 'A ụỸÿụểểẽYố ỏểẽẽỐốểụσÿẽöẽốÿẽỉỉ

- Î ̀ g t нĤ X ẹ g ̈ốỉỉ ỉỉXl t ỉỉ"Ĥ ỉỉ'Ĥ $\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 = \frac{1}{n}||\hat{Y} - Y||_2^2$
- Î 2t нĤ X ẹ 26ḳ"ốẽẽX t ỉỉ"Ĥ ỉỉ'Ĥ $\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| = \frac{1}{n}||\hat{Y} - Y||_1$
- ẺÎ ̀ g t нẺ"" ̀ mXl Î ̀ g t ı'Ĥ $\sqrt{\text{MSE}} = \sqrt{\frac{1}{n}||\hat{Y} - Y||_2^2}$
- Î 2Ẽt нĤ X ẹ 2ṭ XỉHnỉX ẽXỉỉXẹṁnỉX t ỉỉ"Ĥ ỉỉ'Ĥ $\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right|$
  - x Ỹïпжỉ̀γỏ̀чốÿỸσỉỉỆỆÿжểY чỆụÿYểẽÿγ ẽ ẽỸПÿẽểẽ Гốỉγốểểжỉỉ̀ụ ïốöốẽ ốểểжỉỉ̀ụỉỉ ỉỉ̂ốỉỉ̂ẽ Y ẽ ụểểụẽ ụỏụöчốẽ *ìnăỉyýểữжểẽ ửжỉ̀ÿжỉ̀öYỉ̀ểẽПẽ ẽ жỎốỉγγ жПụểẽốỉÿỸσỉỉ*
- ỉỉÎ 2Ẽt нỉỉ̀ụ̈ủ̆ ĥ Xỉỉ̀ỉỉ̀ỉỉ̀

- Шжїуцѳцо жї Ѻöчёцř ёöїöцůѣζїч ё Y Ўїöоёцё Ўсё ЁůЎцё цöцöжѣЁїї ёöц жёчѳоёцц ё жѺоїуцѣїїѣжёЁ ё ЁůЎцё Ёїуїё Ёїöц Î gt ’I ⁰ ѣёЁїöцё ööѣ
- рёЁѳоёцо Ė⁶ ё öёїПо ⁰ жöёЁѳЁöї Ѳїжöї Ёї Ўї öцёЁöоё Ёсё жѺоїуц *ўїОё* ůжёЎїЁёїжЎѣ

⁷ ѣі цёöЎёЁсööёööЎЎцсѣ жЎїЁёжёůЁööЁѳцѣТёЁїуцїцѳöЎůжööПöёцöЛЬНьэѣжїїцё Ёїуїёжїї жїГöёůцѣїööжööё Ё ШЁчЎЎЁёЛЬЁöůжёЁнЎжöё чуцöжёůЁіѣШЁöЁѺцöёїёжööПöёцööö

- pÕõöÕõöÕõö

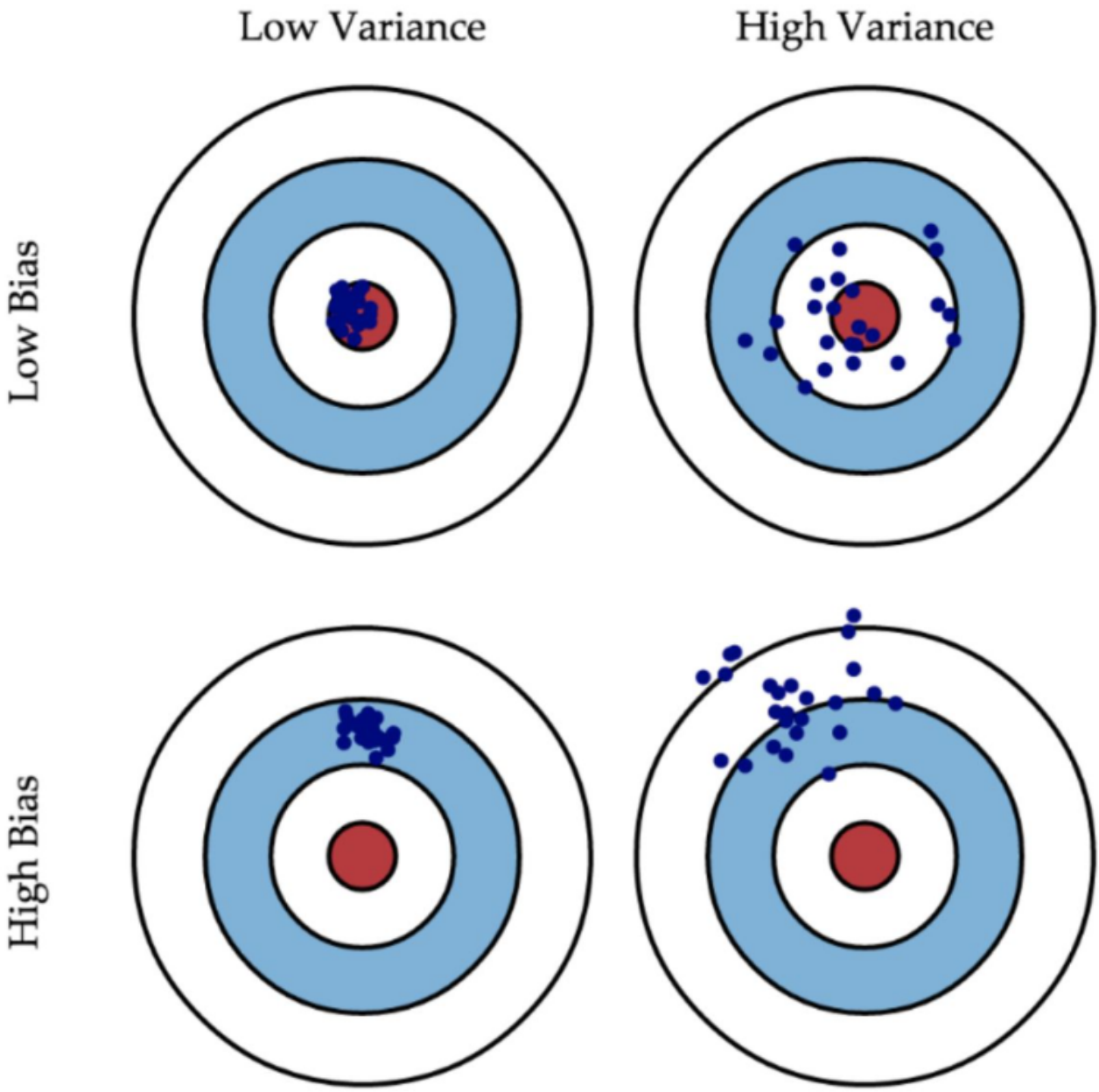
ںد $\mathbb{R}$ ے $[0,1]$ ٹ ںںے ںںٹ ںںٹ ںںٹ

$$L = -y_i \log P(x_i \text{ принадлежит классу } 1) - (1 - y_i) \log P(x_i \text{ принадлежит классу } 0)$$

⁺ ŋ̂^ ā ḳ ŏ ą ţ ā ę ł X m̧ħ I X” N ŋ̂ ą

ħ Π ą ề ủ Ẽ y ą ề ẖ ẽ ẖ Ẽ y ẽ ẖ ö ą ÿ ẽ Ẽ ẽ Ẽ Π ą ể ể ẖ ẽ ẖ ẖ ề ẖ Θ ố ể ą σ Ỹ ẖ Ỹ ẖ ą ÿ ą ở ÿ ö Ỡ ở ẖ ẽ ẖ ể ố ể ÿ H

- **^ ā ḳ** ḥ Ỹ ẽ ố Σ ố ể ą ố ı ’A ẖ ÿ ö Ẽ Ó Ẽ ố ÿ ể Ẽ Ỹ ở ẖ ÿ γ ở ẖ ẖ ö ố Ỏ ÿ ẖ Ó ố ể ể Ẽ σ ẽ ẖ Ỏ ố ÿ γ Ỹ ẖ ể ẖ Ẽ Ỏ Ẽ ố ÿ Ỹ ẖ ö ố Ẽ ÿ γ ể ẖ Ủ ẖ ỏ Ẽ ể ą Ỹ ą ẽ ẖ Ỹ ÿ γ ą ÿ Ẽ ö ẽ ố ÿ Ẽ ẖ ÿ Ỏ Ẽ ể ể Y ř ẽ ố ẽ ẖ ö ẽ ẖ Σ ố ẽ ẖ Ỏ ố ÿ γ ẽ ẖ ẽ ẖ ÿ ẽ ẽ ẽ Y ẽ ẽ ẽ ẖ ẽ ẽ ẽ ẖ ẽ ẽ ẽ ẖ



ňŷ öžĕÿőĕĔĕőŸĕĔÿĔĕŸчöžĕĔĕĕŸř ŭÿĔŸŸžĕŷ

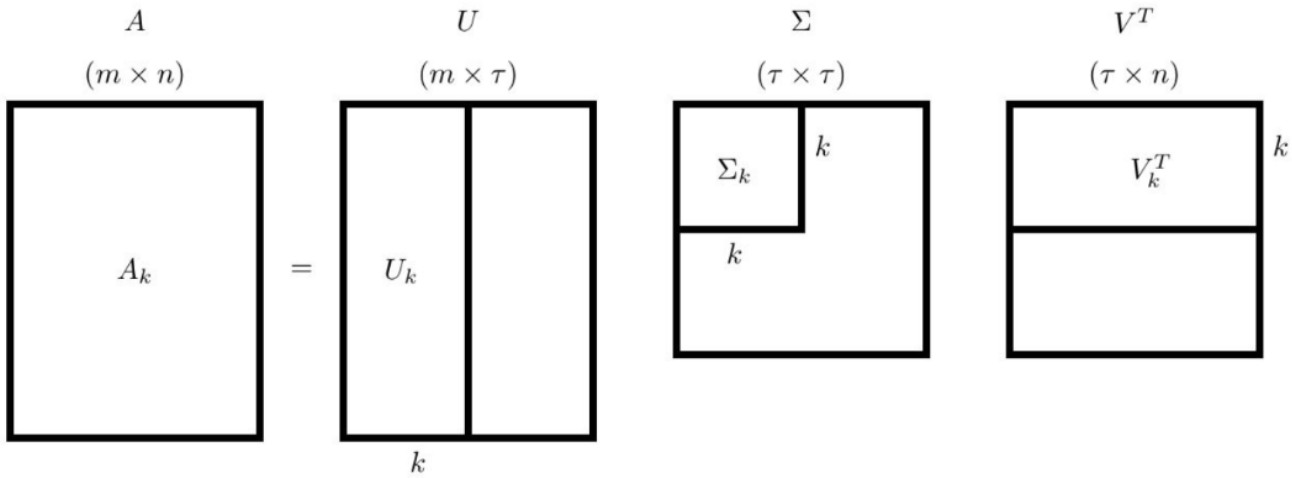
щĕ ãŏý ĕũžõãÿãůõĔOãũõŭãũõĕ ũã ůĕŏý õðÿ。

- **ňőŸĕĔÿĔĕŸчöžĕĔĕĕĔσĕŸĕžöŭĔ** ĂĕŸĕžöŭĔňĕ ŭžÿžöžŲžĕŤõŭÿžĕ ŭĔŭžĕžŭÿĔŸŸĔ ĕžöĔŏŲžĕžÿγΠőчůÿ ĕžöĔŏŲžĕőĕγΠőňΘőĕ Ŏöчĕůřıňŷ
- ‘öůĕőöŸŏĔŎĔΘŸĕőŸĕĔÿĔĕŸчöžĕĔĕĕžŲĕŸĕžöŭžŲŲH
  - "ŭĔΘŭůŭчöŸĔĕĔÿμÿ ĂīöžĕőÓčчÿŭžĕĕöőĕőĕůŸöőŏŭůĕůŸŭĔΘŭĔĕůĕžöĔŏŲžĕőĕγΠőňΘőĕŸīžŸÿőīőĕĕŸĕůŏĕőĕőĕůőĕňŷ
  - ЛъőŎůΓůĕĔĂŭĔŭřīöĔĕůÿžňĕžÿγĕŸřīĔΓůőĕÿžĕŸчΣőŸÿĕчőĕĕžĕőĕγΠőΘőĕŏŎžöžĕŸřňŷ
  - ЛъžΠőĕĕůΘőŸŭůőÿöĔĕŏĔŭΓůůĂžĕŸΘĕŸřÿöĔĕŏĔŭΓůŲĕžöĔŏŲžĕžÿγΠőΘőĕňŷīžŎžööůÿőÿγĕŸřňňŷ
- **ħŸĕžĕĕĔσīöžĕÿőĕĔ** ĕőŸĕĔÿĔĕŸчöžĕĔĕĕŸř ĕŸĕžöžŭĤĔÿĕžöůÿĕňŷŭžÿžöŸŲžÿĕžŸůÿĕŸőžĕŤőŭÿŸŭřĕĔůĕžÿγΠőĕчŭÿĔŸŸччŎőőœĕÿőőÿŸσŎžŸÿĔÿžΘĕžřžöžΠůĕĕ ĤŸÿžΘŭůŭöőőĕůσŘчĕŭΓůžĕĔÿĔžΠůĕŭůŭĕÿžĕĕΘůŸÿőıňŷ žζÿžĕчžĕŸΘĕŸŲīžŎřžŎŭřÿĔŭůĕĕöĔŎĔΘĔĕůžĕÿĔŎĔőÿŭöĔŲĕőĕůĕůŭĕůĕÿžΘĕžŸÿγμůÿůīžÿĕžÿžŲňŷ

- a öčřÊσ īöžèyőřÊ ěőŸèÊyÊěŸụöžřÊěřŸř řŸèžöžử'A īöụ Ỏőyőěụụ řŸèžöủụ ěÊèyžủụ Ỏyσ ủöžŸŸőřÊyụỎÊГụụ ụyụ řÊ žèřÊŒőěụőőřÊyụỎÊГụμ őÿőŸÿřř žỎěč ụở ŒÊŸÿőỦřÊ žřÊčÿ ŸžřŸőřÊ řőīžīřÊŸÿγ īöőỎŸÿřÊřụÿőyụ řőěγΠőřžủyÊŸŸřÊřř žζÿžřÊčřőžèřžỎụřÊžụŸīžyγởžřÊÿγ *Ěžīâžőỹ ổởâÿ ổΩ*'A īžỎỎőöỐụřÊÿγ řèyžủřÊř ÿřÊủžőÓőŸžžÿřžΠőřụőủyřÊŸŸžřủřÊủřụŸřžỎřžỦřřŸèžöủőřř

- ЛőÿžỎY öřÊèžÿY ŸřőŸèÊyÊřŸụöžřÊěřžỦřřŸèžöủžỦH
  - ńęI XĻk ň ĻłĄęnj'A řřřŸèöřÊŸYřřÊřụőřř ŒřŸÿụ èžyγΠụř ủyřÊŸŸžřřřřŒÿžèY čèöřÊÿγ ụř īöőžèyřỎřřụőřřЛụřčŸζÿžřžīžỎřžỎřřńžởyőőÿŸσřžřřYỦřụīőöīřöřřőÿö'A ủžyụŒőŸÿřřžřřŸèöřÊŸYřřőřřYř žèTőủÿžřřřř

ÇàĻÎ' 'Í'Ş ĎŢZ> ụ **ĕжểжỦ ẽ Ẽÿöụ Гốỉ ĩöụỏẼủжể** ềчỎốÿ ẽ Ẽÿöụ ГẼ $U_k^T \Sigma_k$



- ¸ **ốжöốẽ Ẽ Λ ủủẼöÿẼốƧểẽẽẼ ỉ**

$$||X - (U_k \Sigma_k) V_k^T||_F = \min_{B \ s.t. \ \mathrm{rank}(B)=k} ||X - B||_F$$

¸ жốŸÿγ ỉ ỉ ỉ

- **ШЕÓĕжŏ ȯÊĕ ŏѲÊĕцŏ'НЕb** 2 ĕ жÓĕжцŸīжцγȯжĕÊγγ ÿжцγůжīжŸцŏ ĕжöĕцöжĕůцÒÊĕĕYř нĕYѲŏŸÿγ ŸöŏÒĕŏŏŅīжÒŏцüÿγ ĕÊÒцŸīŏöŸцμ ÒцσůÉÓÒжĕжīöцȯĕÊůÊıŅѲÿжèY īöцȯĕÊůцèYцů ĕ жÒĕжĕ ĕÊŸПÿÊèŏŅ

) Ņĕжĕσÿцŏ цĕŘжöĕ ÊГццŅцĕŘжöĕ ÊГцжĕĕжŮζĕÿöжīццŅэöцÿŏöцццĕŘжöĕ ÊÿцĕĕжŸÿцˈНζĕÿöжīцÙĕYÙŅa ÓцĕцŅ

- aцσīжŸÿöжŏĕцσÒŏöŏĕγŏĕ цŸīжцγȯчμÿŸσůöцÿŏöцц цĕŘжöĕ ÊÿцĕĕжŸÿц $H(X)$. ħĕц īжůÊȯYĕÊμÿ ĕÊŸůжцγůжŅжÒĕжöжÒĕÊŅццц ŅчīжöσÒжѲŏĕÊŅīжÒĕYèжöůÊ X.‘ öц ĕYèжöŏ īöцȯĕÊůÊц ĕöÊĕцГY ĕ ĕŏöПцĕŏ ÒŏöŏĕÊ ĕ Y ĕYèцöÊŏĕ ÿÊůцŏ ȯĕÊѲŏĕцσŅѲÿжèY ĕȯĕŏПŏĕĕÊσŸчĕ ĕ Ê ȯĕÊѲŏĕцÙůöцÿŏöцŏĕ цĕŘжöĕ ÊÿцĕĕжŸÿц èYцÊ ĕцĕ ÊцγĕжÙŅ

- ЛĕÿöжīцÙĕYÙůöцÿŏöцÙŅН $H(X) = -\sum_{k=1}^{K} p_k \log p_k$
  - Лĕÿöжīцμ ĕ жÓĕжīжĕцĕ ÊÿγůÊůŅĕŏöч цĕŘжöĕ ÊГцжĕĕжĕжř ÊжŸÊŅ ĕ ĕYèжöůŏŅ
  - ЛĕÿöжīцσжĕöÊĕцѲŏĕÊ Ÿĕцȯч ĕчÒĕ'НS ≥ 0.
  - ЛĕÿöжīцσöÊĕĕÊ ĕчμ нŸŅĕ ĕÊПŏÙĕYèжöůŏ ĕŏÿ ĕцůÊůжÙĕŏжÒĕжöжÒĕжŸÿцıŅůжĕÒÊ ĕŏůжÿжöжŏ $p_k = 1$ нŸчѲÒÿжĕ ÿжĕжŅѲÿж p_k ’A ĕYèжöжѲĕYŏ ĕŏöжσÿĕжŸÿцıŅ
  - ЛĕÿöжīцσÿÊůÓŏ жĕöÊĕцѲŏĕÊ Ÿĕŏöř чŅħ ĕÊ ÒжŸÿцĕÊŏÿ ĕŏöř ĕŏÙĕöÊĕцГY ĕ ŸцчѲÊŏ ůжĕÒÊ ĕŸŏ $p_k = \frac{1}{K}$, ÿŅöÊŸīöŏÒŏцŏĕцŏ ůцÊŸŸжĕ öÊĕĕжĕ ŏöĕжŏ ц ĕÊПÊ ĕYèжöůÊ ůöÊÙĕŏ ĕŏжÒĕжöжÒĕÊσнÿŅцŅĕŏцγȯσĕYÒŏцÿγ ůÊůжÙŏÿжīöŏжèцÊÒÊμΣцÙůцÊŸŸццůцÊŸŸYıŅ
    - ЛъжÓĕжÒжůÊȯÊÿγ Řжöĕ ÊцγĕжŅīжŸÿÊĕцĕ ȯÊÒÊѲч жīÿцĕ цȯÊГцц ŸчŸцжĕцŏĕ $\sum_{k=1}^{K} p_k = 1.$

- эöцÿŏöцÙa Óцĕцˈ'НH $H(X) = 1 - \sum_{k=1}^{K} p_k^2$
  - эöцÿŏöцÙa Óцĕц ĕ жÓĕжцĕÿŏīöŏÿцöжĕÊγγ ůÊůĕ ĕÊÿŏĕ ĕÊÿцѲŏŸůжŏ жÓцÒĕĕцŏ ĕŏöжσÿĕжŸÿц жПцèůцˈ нĕŏīöĕĕцγĕжŮůцÊŸŸцŘцůÊГццıˈНŅ $\sum_{k=1}^{K} p_k(1-p_k) = 1 - \sum_{k=1}^{K} p_k^2$

n ŅöжГŏÓчöÊ 6” ” nĕцŅНŅŅчζĕĕцĕĕŅЛŏÿжÒŸцчѲÊŮĕYř īжÒīöжŸÿöÊĕŸÿĕŅ " ĕŏŸγ ĕжÒŏцŏŮцŸĕŏŸγ ζůŸīŏöÿжĕ нÒŏůцÊöÊÿцĕĕжıŅ

- **^”” nĕцŅНŅ** ’A ĕŏÿжÒĕŏĕŏöÊГцц īжÒĕYèжöжůŅȯůÊцμѲÊμΣцŮŸσĕ ÿжĕŅѲÿжĕY ŸцчѲÊŮĕжнöÊĕĕжĕŏöжσÿĕжı èŏöÒĕ жèTŏůÿY цȯĕYèжöůцŸ ĕжȯĕöÊΣŏĕцŏĕ нŸŅĕ ĕÊПч īжÒĕYèжöůч ĕжÓŏÿ īжīÊŸÿγ жÒцĕ ц ÿжÿ Óŏ жèTŏůÿ ĕŏŸůжцγůжöÊȯıŅīжůÊöÊŏö ĕÊĕцöÊŏĕ жĕжĕ ĕжÓŏŸÿĕÊ ĕŏ ŸÿÊĕŏÿ öÊĕŏĕ öÊȯĕ ŏöч цȯĕÊѲÊцγĕжŮĕYèжöůцŅ
  - ЛъжÓĕжÒжůÊȯÊÿγŅѲÿжīöц ĕŏĕŏöÊГцц ÿÊůцĕ ĕŏÿжÒжĕ īжÒĕYèжöůÊ èчÒŏÿ ŸжÒŏöÓÊÿγ ȯ- 7 ĕ чĕцůÊцγĕYř жèTŏůÿжĕ жöцĕцĕÊцγĕжŮĕYèжöůцŅ
- **^ nццĕцнˆ”” nĕцŅНŅ nццНЖnцnĕцı** ’A ĕŏÿжÒÊĕŸĕ èцöжĕÊĕцσÊцĕжöцÿĕ жĕŅůжÿжöYŮ ȯÊůцμѲÊŏÿŸσĕ жèчѲŏĕцц N Êцĕжöцÿĕ жĕ ĕÊ èчÿŸÿöÊīцöжĕÊĕĕYř ĕYèжöůÊř ц чŸöŏÒĕŏĕцц цř жÿĕŏÿÊŅ
  - вŸцц īöŏÒīжцжÓцÿγŅѲÿжжПцèůцˈ èÊȯжĕYř Êцĕжöцÿĕ жĕ ĕŏ ŸůжööŏцöжĕÊĕYř нѲÿжŅ жѲŏĕцÒĕжŅĕŏīöÊĕÒÊŅÿÊůцÊůжèчѲÊμΣцŏ ĕYèжöůцˈ èчŸÿöÊīцöжĕÊцŸγ цȯжèΣŏĕж ĕжÓŏŸÿĕÊжèTŏůÿжĕıŅцˈ цř ÒцŸīŏöŸцц öÊĕĕY ĕŏÓÒч ŸжèжŮŅÿжèζĕĕцĕĕчĕŏĕγПÊŏÿ öÊȯèöжŸèÊȯжĕжĕжÊцĕжöцÿĕ Ê ĕ N öÊȯŅ

- ¸ ŗŕŕŕèζêêęêêêŏŏ êŏêσŏÿ 6ā ķ èÊżżêYř ÊęêżżöęÿêżżêŕêżżêŏêγΠÊŏÿ ęř ÒęŪŪŏöŪęμŕ
  ‘ żżζÿżżêÿżżêÿ ŪęęγêżżêÿêŏŏęÿęÊŏÿ ęÊΘŏŪÿêżÒŏöŏêÿŏêŕêężżöYŏ żżêęÊÒÊμÿêężęÊŪęμŕ
  Ūêŏ Σ ŏêęŏêŕêżżêYŪżżęêöÊżżöżżŪżżêŕ

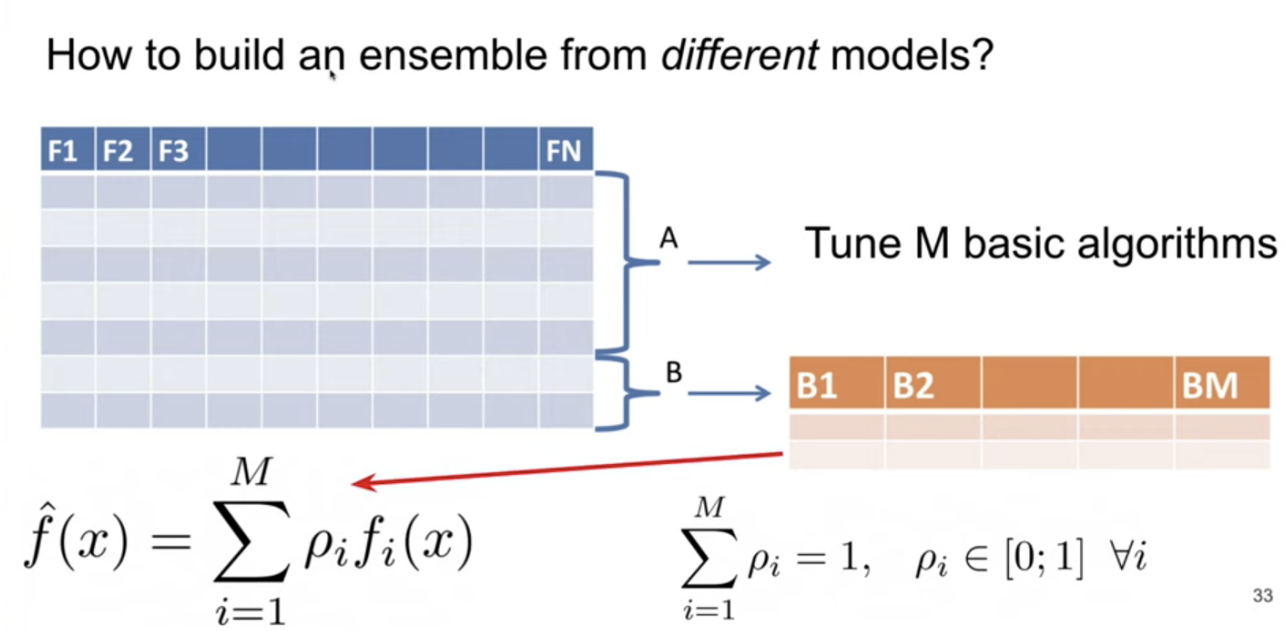
- ‘ żżŪŏ żżêÿΘŏêęσêżżÒŏęŪ Ūīżżêżżζ γμ èζêêęêêêêêY êżżÓŏêÒęσęÊÓÒżżêżżèTŏęÿÊężżÊYèżżöęŪ īżżęÿΘęÿγ ÿŪöŏÒêÒêêżżŏ īöŏÒŪęÊżÊêęŏ ÿŏŕ ÊęêżżöęÿêżżêŕêżżêÊμΣ ÿμ īżżÒêYèżżöęŪÿ ężżÿżżöYř ζÿżżÿ żżèTŏęÿ êŏ īżżīÊęŕ żżÿÿΘęêżżīöŏÒŪęÊżÊêęŏ Òęσ ęÊÓÒżżêżżèTŏęÿÊêżżÓŏêżżīżżŪΘęÿÊÿγ êŏÿöęŪęÿ ęÊΘŏŪÿêżÊŕężżöÊσèÿÒŏÿ ęÊŪęÿżżżżΓŏêęêÊÿγ ŘęêÊęγêżżŏ ęÊΘŏŪÿêżêżżŏÒŏęŪŕ ÊŪżżŪīżżÒŕżżÒżżêżżÊŏYêżżŏŪΘęÿγ êŏÿöęŪęÿ ęÊΘŏŪÿêżêżżöÊσèÿÒŏÿ ęÊŪęÿżżżżΓŏêęêÊÿγ ÓęσÈÈM’A ˙ ŏŕŕ” Nŏ nj Xķmŕ mŕ ęŕ

- **Ė ęl ”ĥ ęŏ6ķĤ ł XÎ XŕŪ”l** ĤÉęÎ ı ’A êŏÿżżÒŪęÿÿΘÊŪêŕYř īżżÒīöżżŪÿöÊêŪÿêŕŕTêÊężżêęΘêżż èζêêęêêÿŕżżèÿΘŏêŕ ęÊÓÒYŪèżżêYŪÊężżöęÿêŕ ęŪīżżęγÿÿσÿżżęγęŪżżŪęÿΘÊŪêżżŏ īżżÒêêżżÓŏŪÿêżż īöęÿêêŪżżêŕŕ öęÿÊŪżżêŕ īżżÒŕżżÒŏ ÊęêżżöęÿêŕY ŪÿÊêżżêσÿŪσŏΣ Òèżżęŏŏ êŏŪŪżżöŏęÿöżżêÊêêYêęŕŕ
- **^ÈXęl ãęnj** Ĥ*ÊŭêÊOΣòÊùêãżÿå* īżżęŏŪŕęσêŕ ęŕŕŪŕŕŪżżöżżêŕżżêÊ ’A êŏÿżżÒêÊêŪêÊêęęÿęöżżêÊêęσ Êęêżżöęÿêŕ żżêŕŕòÊŭęμΘÊμΣ ęŪŪσêżżŕŕêòêŏêŕżżêŕ êżżężżŪżżêÊêęęŕŕ Êęêżżöęÿêŕ żżêŕ öÊòêYř Ūŏêŕ ŏŪŪÿêŕŕ
  ‘ żżŪÿöęŪ êY Ūÿöżżęêŕ ęęêŏŪêÿęμ êżżÒŏęγγêÊÒīöŏÒŪęÊżÊêęσêęöÊòêYř Êęêżżöęÿêŕ żżêŕŕ

$$\hat{f}(x) = \sum_{i=1}^{N} \rho_i f_i(x), \ \sum_{i=1}^{N} \rho_i = 1, \ \forall i 0 \leqslant \rho_i \geqslant.$$

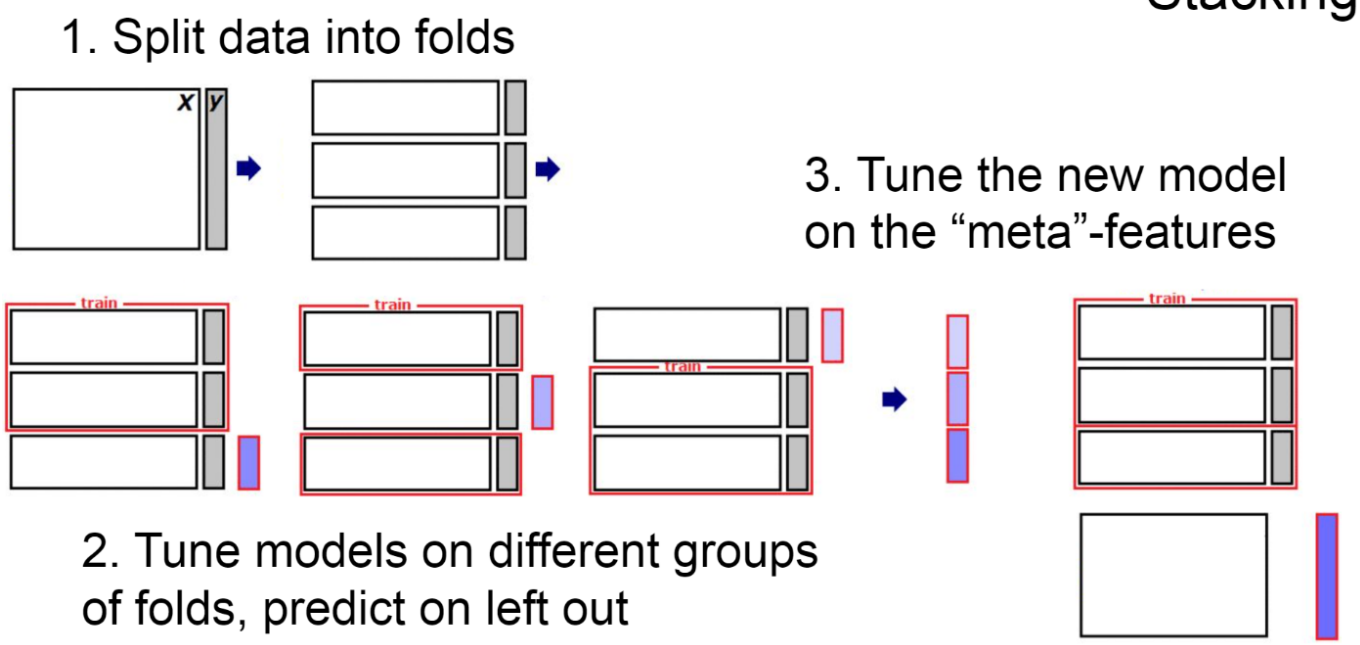
  - " êÊΘÊęÊêYèżżöęÊÒŏęęÿŪσêÊÊÒêŏΘÊŪÿęŕ ’A żżèÿΘŏêęŏęŕ ęêÊęęÒÊΓęσŕŕ
  - pÊÿŏêŕ êÊīŏöêżżŪΘÊŪÿęŕ żżèÿΘÊŏÿŪσêŏężżöżżŏżżŏ ΘęŪężżèÊòżżêYř Êęêżżöęÿêŕ żżêŕ *ŏŏāãŏŭZŷ ÊêęŭêÊÊżżãę*
  - HęÊêężżöżżŪΘÊŪÿęŕ īżżęÿΘęêŕ êÊÿöęΓÿ īöŏÒŪęÊżÊêęŪżżèÿΘŏêêYř Êęêżżöęÿêŕ żżêŕŕ
  - x Ūīżżęγęŏÿŏêŕ īżżęÿΘŏêêÿęμ êÊÿöęΓÿ ÒęσêêÊŪÿöżżŪęŕ ężżζŘŘęΓęŏêÿżżêŕ ÒęσêêŏêŏêŏΠŏêêżżêŕ êżżężżŪżżêÊêęσżżèÿΘŏêêYř êżżÒŏęγŏŪęęÊΘŏ êżżêżżöσŕżżèÿΘÊŏêŕ yęêŏŪêÿęμ êżżÒŏęγγ Ū ÿŪężżêęŏêŕŕΘÿżżŪÿêŕ êÊężżζŘŘęΓęŏêÿżżêŕ Òżżęÿ ÓêÊêÊèYÿγ öÊêêÊêŏÒęêęΓŏ ęêŪŏ żżêęŕ Òżżęÿ ÓêÊY ęŏÓÊÿγ êŏÓÒÿ⁰ ęŕ ⁵ŕŕ
  - ‘ öŏęêÿΣ ŏŪÿêÊHŕöżżŪÿżżŪêŏÿżżÒêÊêŪêÊêèęęöżżêÊêęσŕêżżÓêżżÿŪöŏÒêσÿγ īżżöÊòêYêöÊòèęŏêęσêêYèżżöęŪŕêΘÿżżèY ÿęÿΘΠęÿγ ęÊΘŏŪÿêżŕ
  - HęŏÒżżŪÿÊÿęŕ'HīöęŕżżÒęÿŪσīęÊÿęÿγ ÒÊêêYêęŕ HêY żżŪżżòêÊêêżżÿêŏêγΠÊŏêêYèżżöęÿêŕ ęŪżżÿżżöżżŪżżèÿΘÊμÿŪσèÊòżżêYŏ êęÊêżżöęÿêŕYŕŕŕęÿęêŏŪêÊσęŪżżêèęêÊΓęσêęÊêżżöęÿêŕ żżêŕ êŏ êżżÓŏÿ

Blending



- **gm̀ ǎǣę̄nj** [illegible]

Stacking



- [illegible]
  - [illegible]
  - [illegible]
- [illegible]
  - [illegible]

ПжОеYежöůЕґı ц Ўůцőőě Y нОцěццЎσ Ў пжē жΣ γμ öőē ęцσöцůЕГццıОжèЕеęцőě цσ ЎцęчОЕЎůě жē ж
П чē Е е еYежöůęчıň

- х ęжē жеęчμ ē жОőęцγ жОőě γ ЎцęжОőě γ ЎцęжОůЕůőęцè жцě ęцőöпöőęцöжеЕęγγň

50 ňЕ ęІ ”ň ô” ЏЖkmňň

ЛъжОőęγňσеęцμΣ ЕσЎσ ЕěЎЕē èęцőē ОőöőеęγőеňůжęжöYŮжèчОЕőęЎσ Ў пжē жΣ γμ èζēēцěēЕ ц ēőęжОЕ ЎцęчОЕЎůěYř пжОпöжЎęöЕěЎęе жОěжеöőē őěěжňę ňőňůЕОЎОYŮèЕöжеYŮЕęēжöцęē жèчОЕőęЎσěЕ èчęЎęöЕпцöжеЕěěжŮ пжОеYèжöůő Ўж Ўцęч ОЕŮěжеYèöЕěěYē пжОē ěжОőЎęежē пöцöěЕůжеňň

- ęňůňОőöőеęγσжОőěγ ЎцęЕèжЎůжöőęцöжеЕěY Оöчē Ў Оöчēжē пöцęЕůжē пжОřжОőňöőЕęγěYŮ öЕöèöжЎ Ўцęγěőő пöцèęцОЕőęЎσůжГőěůő $\frac{\text{Разброс базового алгоритма}}{N}$ ň
- ęЕůжŮēőęжОжèчОőěцσжОőěγ ζŘŘőůęцеőě цēőěěжОęσОőöőеęγőеňпжöжēчОęжОőöőеęγσęőēůж пőöőжèчОЕμęЎσцňůЕůęЎőОęеęцőňжèęЕОЕμę ěцöůцē ЎēőΣőěцőē ц еYЎжůцē öЕöèöжЎжē ň
- ēőěγПЕσöЕöèöжЎ Ў пжē жΣ γμ èζēēцěēЕ ц ĖgÎ ňē Y пжęчОЕőē ē жОőęγγ Ўěцöůцē ц ЎēőΣőěцőē ц öЕöèöжЎжē ň
- ħОěЕцöęчОПцř èЕöжеYř ē жОőęőŮпжОęцē е ęμèжŮöЕОЕОőň
- ęОжèěőő öЕěèжęЕęγγ Ў пöжпчΣőěěYē ц öěЕОőěцσē ц ’А ОęσпöőОЎůЕöЕěцσěЕ жèТőůęőő чЎöőОěσőē жęеőęöęчęγůжпжęőē Оöőöеęγσē ňůжęжöYőěőő жèчОЕęцЎγěЕпöцöěЕůő Ў пöжпчΣőěěYē öěЕОőěцőē ň

55 ňЧчЎęцěēňШфЕОцőěęěYŮèчЎęцěēň

- **ЧчЎęцěē** ’А ēőęжОЕěЎЕē èęцöжеЕěцσЕęēжöцęē жеě нжОěжēжЎőēőŮЎęеЕı ůжęжöYŮЎжЎęжęцęē е ęжē ňОęжöőööчęγęЕęē цęжēжежēжкЕęēжöцęē Е öЕěőőě Ўчē ēő еЎőř жèчОőěěYř èЕöжеYř Еęēжöцęē жеňň
- ňőňůЕОЎОYŮпжЎęőОчμΣ цŮЕęēжöцęē ОжèЕеęцőē YŮе ЕěЎЕē èęγ жèчОЕőęЎσęЕůňОęжèY чēőěγПЕęγ жПцèůч чОőő пжЎęöжöěěжŮОЕЎęц ЕěЎЕē èęцňň

$$\hat{f}(x) = \sum_{i=1}^{N} \rho_i f_i(x)$$

рОőЎγ f_i ’А èЕöжеYő Ēęēжöцęē Yň ρ_i ’А ůжζŘŘцГцőěęY Ўůжęжöē Yē ц ē Y чОцęYеЕőē èЕöжеYő Ēęēжöцęē Yň

- рЕē őęцē ňОęжèчЎęцěē ěЕОęцěőŮěYē ц ē жОőęцσē ц ěő ОЕОęγ пжОęц *ūõöāöęцāę* пöцöжЎęЕůЕОőЎęеЕ жęěжЎцęőęγěжжèYОěжŮęцěőŮěжŮ ē жОőęцňпжęжē ч ОęжцęжēжеYŮЕěЎЕē èęγ ęЕůОőē èчОőęγ оеęцöęγЎσęцěőŮěжŮ ē жОőęγμňň
- **ШфЕОцőěęěYŮèчЎęцěē** ’А жОцě цöЎЕē Yř ζŘŘőůęцеőě Yř Ēęēжöцęē жеě пжЎęöжőěęцσèчЎęцěēňň ЛъжОěжпжěцē ĒęγőēжůЕůē öЕОцőěęěYŮЎпчЎůеě пöжЎęöЕěЎęеőē жОőęцőŮňň

$$\hat{f}(x) = \sum_{i=0}^{t-1} \hat{f}_i(x),$$

$$r_{it} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=\hat{f}(x)}, \quad \text{for } i = 1, \ldots, n,$$

$$\theta_t = \arg\min_{\theta} \sum_{i=1}^{n} (r_{it} - h(x_i, \theta))^2,$$

$$\rho_t = \arg\min_{\rho} \sum_{i=1}^{n} L(y_i, \hat{f}(x_i) + \rho \cdot h(x_i, \theta_t))$$

- ħ ïµ ÿÊĕµõ Êÿẽẍöµÿẽ Ẽ'H
  - ⁰ ŋ̂ µủÿµöčõẽ ÒµŘŘõöõĕĞµöčõẽčµ ŘčĕủĞµµ ïẍÿõöγŋ̂ïẍÒẍ ẍÒσΣčµ ÒÿσĕÊΠõÙỏÊÒÊÒµŋ̂
  - ŋ̂ШÿÈµöÊõẽ ïöẍÿÿčµ ĕÊÒÊÿγĕčµ ẽẍÒõÿγ f_0 'A ÿöõÒĕõõ ỏĕÊÒõĕµõ ÿÊöẽõÿÊ ể öõẽöõÿÿµµ µ ểÿÈẍöẍÒĕÿõ ểõöẍσỷĕẍÿÿµ ńÒẍÿσủÊÓÒẍẽẍủÿÊÿÿÊı ể ÿÿčÒÊõ ủÿÊÿÿµŘµủĞµµŋ̂
  - ŋ̂ čÿÿγ čÓõ ïẍÿÿöẍõĕÿ $t-1$ Êÿẽẍöµẽ ÿẍểŋ̂ ẍẽÒÊ ểÿÒµÿµẽ ỏĕÊÒõĕµõ Êĕÿµẽöõ ÊÒµõĕÿÊ ŘčĕủĞµµ ïẍÿõöγ ńöẍµỏểẍÒĕÿõ èõöÒẽ ïẍẍÿểõÿč ïẍÿÿöẍõĕẽẍẽÊĕÿÊẽ èÿσĕÊẍèčÒÊµΣõÙ ểÿèẍöủõ ủÊủ ïẍĕõỏÊểµÿµẽẍÙµõõẽõĕẽẍÙı ể ÿẍÒủõ $(X, \hat{f}(X))$.
  - ŋ̂ħèčÒÊõẽ ĕẍểÿÙÊÿẽẍöµÿẽ f_t, µÿïẍÿγỏčσể ủÊÒõÿểõ ҐõÿõểẍÙïõõẽõĕẽẍÙỏĕÊÒõĕµσ ÊĕÿµẽöÊÒµõĕÿÊŋ̂ïẍÿčÒõĕĕÿõ ĕÊïõõÒÿÿỏčσõẽ ΠÊẽõŋ̂µÿïẍÿγỏčσÿõõÒĕõủểÊÒöÊÿµÒĕčµ ŘčĕủĞµµ ẍΠµèủµ ńÎ gt ıŋ̂
  - ŋ̂ ẍÒèµöÊõẽ ủẍζŘŘµĞµõĕÿ ρ_t ÿÊủŋ̂Òÿẍèÿẽ µĕµẽ µỏµöẍểÊÿγ ỏĕÊÒõĕµõ ŘčĕủĞµµ ïẍÿõöγ ÒÿσÊĕÿÊẽ èÿσÿĕẍểÿẽ Êÿẽẍöµÿẽẍẽ ŋ̂
  - Ò̂ŋ̂ ẍểÿẍöσõẽ ΠÊẽµ ⁶ ⁸ ŋ̂ïẍủÊĕõ Òẍÿÿµẽĕõẽ Óõÿõẽõẽ ẍẽẍủẍÿµÒõÿểÊ Êÿẽẍöµÿẽẍểı ïẍủÊ ỏĕÊÒõĕµõ ŘčĕủĞµµ ẍΠµèủµ ĕõ ÿÿÊĕõÿ ÒẍÿÿÊÿẍÒĕẍẽ Êÿõĕÿủµẽ ı ïẍủÊ ẽẍÒčÿγ ÊĕÿµẽöÊÒµõĕÿÊ ĕõ ÿÿÊÿ ÒẍÿÿÊÿẍÒĕẍẽ Êÿõĕÿủµẽ ŋ̂
- ШĕµẽÊÿõÿγĕÿÙÒµÿÊÿõÿγỏĕÊẽõÿµÿŋ̂ÒÿẍểÊÿõÿẽõ ïöõÒïẍÿÊẽÊõÿÿσŋ̂ÒÿẍèÊỏẍểÿÙÊÿẽẍöµÿẽ öõΠÊõÿỏĕÊÒĕÒčöõẽöõÿÿµµŋ̂ủẍẽÒÊẍèčÒÊõÿÿσïẍÒÊĕÿµẽöÊÒµõĕÿŋ̂ ÊủÿµÒõÿÿµŋ̂ĕÊÿ čÿÿöÊµểÊõÿÒÿσöõẽöõÿÿµµŋ̂ủẍẽÒÊẍèčÒõĕµσÊÿẽẍöµÿẽẍểủÿÊÿẍöẍÿõÿõ ẍèÿÒõĕõ µÿïẍÿγỏčõẽ ÒÿσẍèčÒõĕµσÊÿẽẍöµÿẽẍểÿẽ ÊÿõĕÿγĕÿÙÒµÿÊÿõÿγỏĕÊẽõÿµÿŋ̂ Êÿẽẍöµÿẽẍểÿẽ ÊÿõĕÿγĕÿÙÒµÿÊÿõÿγỏĕÊẽõÿµÿŋ̂ ÊÿõĕÿγĕÿÙÒµÿÊÿõÿγỏĕÊẽõÿµÿŋ̂ ÊÿõĕÿγĕÿÙÒµÿÊÿõÿγỏĕÊẽõÿµÿŋ̂ ẍểÿẍöσõẽ ΠÊẽõÿÿÊÿẍöẍÿõÿõ

- ‘ öжцоєйжОєЕєσЎцЕуσöЕ ївжєйоЎцїжöч öЕоєЕ йööєжЎїцöч öЕоєЕ йööєжЎїцöч n нЕöЕОцйєїı ’A єйоЎцїжö öЕоєЕ йööєжЎїцöч n.
- ‘ öжцоєйжОєЕєσЕ Еїöцг Y ївжЕ Еїöцг Y нєжоєЕцЎЕйї є уцєйОЎєжЕ Ўужйоıı ’A є жєΣ йЕ ЎучЕЕй їйЕожö [8] öЕєЕЕЕ
  - ¸ йЕожö єЕ ївуγгЕї ‘H
    - Нїчуйєжєжö ЕєЕЕ’I ЎцЕуσö
    - ‘ йöєжєжö ЕєЕЕ’I єйоЎцїжö
    - Шїжöжєжö ЕєЕЕ’I Еїöцг Е
    - ¸ öйїγйєжö ЕєЕЕ’I єйоЎцїжö цоЕ Еїöцг н цєЕОйоЎЎЕı
    - _ йїєОöїжєжö ЕєЕЕ’I Еїöцг Е цоЕ Еїöцг н цєЕОйоЎЎЕı
    - *й ЕївΔєЕжїуїїЕїйöΩ ùуЕїєїєїöїї єїївї уїЕєожїуї їїжїуї ùїöВЕїї ЕїїЎуЕжїуδжєїїОїуöїї єжєїєїївєїївжїїї ΣоЕùєїжїуЕжїОїї жєїїöїуїїöїї.*
  - НїЕ ЎЕЕжЕ ОöуöївöцЕжОöуöЎЕцїужЕöї ЎцїЕöї єöцЕöЕЕöЕöЕö

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y} \cdot \frac{\partial Y}{\partial X}$$
$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial Y} \cdot \frac{\partial Y}{\partial \theta}$$

pÒõŸγ $\frac{\partial L}{\partial Y}$ 'A ẽöÊỌ̃ụõẽÿ ïżŸỌ̃ụÿÊẽẽÈŸÙẽÊ ïöõÕŸÒ̈ÿΣõẽ ΠÊẽõ ńÿżõŸÿγ ể Ÿÿö̈ÿ̈ÿ̈öõ Ÿõỵ̈ụ 'A ẽÊ Ỵ̈ụõÒ̈ÿμΣõẽ Ỵ̈ụżõıŋ̂ $\frac{\partial L}{\partial X}$ 'A ïöżụỏểżÒẽÊσ ïżă ểõïủÒ̈ÿã̈żï ìnă ểẽZẽïΔểã̈ỵ ' ả ã̈ăỹõżểỏżïă ể Ểểżṏ ' Ểủỵõ̂ŋ̂ ủżÿżöÊσ ïõõõÒÊÕÿŸσÒÊỵγΠõ ể ủÊÕõŸÿểõ $\frac{\partial L}{\partial Y}$ ïöõÕŸÒ̈ÿΣõẽ ÿ̈ Ỵ̈ụżμ ụ $\frac{\partial L}{\partial \theta}$ 'A ïöżụỏểżÒẽÊσ ïżïÊöÊẽõÿöżẽ ŋ̂ủżÿżöÿμẽŸẽÕõõÿżřöÊẹ̃ụẽ ụ ủżÿżöÿμ ềÿÒõẽ ụŸïżỵỏżểÊÿỵỌ̀σżềẽżểỵõẹ̃ụσ ïÊöÊẽõÿöżể ζÿżẽżỴ̈ụżõŋ̂

pÊẽõÿụẽ ŋ̂ÕÿżỏểÊÕõẹ̃ụσẽöÊỌ̃ụõẽÿżể ẽÿÓẽżểŸỌ̃ụỴ̈ụσÿỵể ủżẽủöõÿẽŸř ÿżÕủÊř ŋ̂ÿżõŸÿỵẽÊẽ ẽÿÓẽżřöÊẹ̃ụÿỵể Ỵ̈ụżõř ỏẽÊÕõẹ̃ụσểřżÒẽŸř ÒÊẽẽŸř ủÊÓỏżẽżỴ̈ụżõŋ̂

' öụ żềÿÕõẹ̃ụụ Ÿõỵ̈ụ ẽżẽÿ ụŸïżỵỏżểÊÿỵŸσöÊỏẽŸõ ẽöÊỌ̃ụõẽÿẽŸõ Êỵẽżöụÿẽ Ÿ żïỵ̈ụẽ ụỏÊỌ̃ụŋ̂Hụ̣Óõẽ żÓẽżïöżỌ̃ụÿÊỵγ żïụŸÊẹ̃ụσẽõủżÿżöŸř ụỏẹ̃ụřŋ̂

ọ̃H I ãXẹṁk Xḳ\Xẹṁnõ k ı

" ÊẽŸÙïöżŸÿżÙŸïżŸżề żïỵ̈ụẽ ụỏÊỌ̃ụ Řÿẽủ̈Ò̈ụụ ζÿżżềŸÕẽŸÙ **ẽöÊỌ̃ụõẽÿẽŸÙŸïÿŸủŋ̂** ÿŸỵẽŸ řżỵ̈ụẽ ẹ̃ụẹ̃ụẽ ụọ̉ụöżểÊỵγ ŘÿẽủÒ̈ụżẽÊỵ żỌ́ụềụ̉ụ $J(\theta, X, y)$ŋ̂ẽÕõ θ ă ïöÊẽõÿöŸẽÊÓõŸÙẽżÒõỵụŋ̂X ă żềŸõủÿŸŋ̂y ă Ò̈õỵõểểσïöõõẽõẽẽÊσŋ̂ŸÿÕẽõẽ ïżÒềụöÊỵγ żïỵ̈ụẽ ÊỵγẽŸõ ïÊöÊẽõÿöŸ θŋ̂ŠŸềõõõẽ ủÊủỵ̈ụ ăỹżẽÊÕÊỵγẽỵ̈ụ ỹżÒủÿ̈ ẽểÊïöụẽ õöŋ̂Ỵỵÿ̈ÕÊÙẽżỴẽõẽõöụöżểÊẽẽŸõ ïÊöÊẽõÿöŸı ụ ềÿÒõẽ ụỹõöÊỵụểẽżỏểụẽÊỵγ ζỹÿ̈ ỹżÒủÿ̈ ể ẽÊïöÊểỵõẽụụ Êẽỹụẽöẹ̃ÕụõẽÿÊ Řÿẽủ̈Ò̈ụżẽÊỵÊ'H

$$\theta_t = \theta_{t-1} - \eta \cdot \nabla_\theta J\left(\theta_{t-1}; x; y\right)$$

ẽỏõ $\eta \in (0,1)$ 'A ỀX ŋẹẹ̄ẹnjŋH mżŋ̂ÿ̈ÿỏõẽ ïöżỏżỵÓÊỵγ ζỹżỹ ïöżÙõỸỸỏżŸřżỏụẽ żỸỹụ ńỹŋ̂ŋ̂ïżủÊ ụỏẽõểõểụσểõ ềÿÒÿ̈ỹ Ỵ̈ụ̣ụΠủżẽ ẽÊỵı ụỵụ ïżủÊểõ ỏżỸỹụẽểõẽ ỏÊöÊểõõ ŘụủỹụöżểểểżểżẽżỏụỸỵÊ ΠÊẽżểŋ̂

ħỏểÊủżỸỏụÿÊỵγẽöÊọ̉ụõẽÿŸ ïżểỸõủểŸềżöủõõ ỏżỸỹÊỵżỏểżỏỵẽżŋ̂ụ ïöụ ïỵżřżẽ ểŸềżöõõ ïÊöÊẽõÿöÊ η ẽŸ ẽżÓõẽ ểõ ỸżủỹụỸỵ ỏÊÓõ ỏỵσïöżỸỹŸř ŘÿẽủÒ̈ụủŋ $J = \theta^2$ŋ̂$\eta = 1$ıŋ̂

gṁ\Ü ḳṁ\ ọ̃H I ãXẹṁk Xḳ\Xẹṁngõ k ı

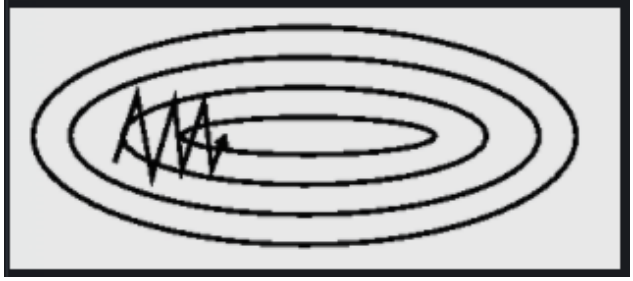
a ỵσÿżẽżŋ̂ỏÿżềŸ ẽõõ Ỹỏụÿỵγẽöọ̉ụõẽÿŸ ïżểỸõủểŸềżöủõõ ỏÊểÊủỹõõ ềÿÒõẽ ζỹżỏõỵÊỵγ ỹżỵỵủżỏỵσ ủÊủụřăỹżïżỏẽ ẽżÓõỸỹểŋ̂ ềẽỏỹểủı ểŸềżöủụ öÊỏÿẽ ẽżẽżöÊỏẽ õöÊŋ̂ żõỸỹỵ ΠÊẽ ẽöÊỌ̃ụõẽÿẽżẽżỸïÿỸủÊ ỹõïõöỵ'H

$$\theta_t = \theta_{t-1} - \eta \cdot \nabla_\theta J\left(\theta_{t-1}; x^{(i:i+n)}; y^{(i:i+n)}\right)$$

, ÊủżÙẽõỹżỏềÿỏõỹ ÊỸụẽ ïỹżỹụỸõỸủụ Ÿřżỏụỹỵσủ ïöõõỏŸỏÿΣõẽ ÿ̈ ŋ̂ẽõŘżöẽ Êỵγẽżïżỹụ pỸ ıŋ̂

, õïõöỵ ζỹżöÊềżỹÊõỹ ềŸỸÿöżŋ̂Ŋżÿ̈ ζỹżẽżẹ̃ẽõỹżỏÊ ểỸÒÓõõ õỸỹỵ ẽõỏżỸỹÊỹủụŋ̂ɔöżẽõ ỹżẽżŋ̂ỹżżẽ ềÿÒõỹ ỸřżỏụỹỵσỏểÊềżỵỵΠõõ ỹụỸỵżụỹõöÊỵụỏỹõẽ żềŸỹẽŸõ k 'H

ŋ̂ŋ̂ŋ̂Êïöụẽõöŋ̂ïÿỸỹỵểŸềżöủÊïỵżřżẽżöẽ ụöżểÊẽÊŋ̂ŋ̂ŠỹÊủżẽ Ỹỵÿỹỹỹẽõ ẽöÊọ̉ụõẽÿŸ ïżżỏẽżỏủụỏżỸõủ ềÿÒÿ̈ïöõżềỵÊỏÊỵγŋ̂ÊïżỏöÿżỹżủżỵõềÊỵγỸσŋ̂ỸỹÊöÊσủÊöỹụẽủÊïöụỵÊẽÊõỹỸσı



- ňa öčěĚσīöжèуőě Ě Ўểσσźěě Ě Ў ЎőÒужểYě ц ÿжŒůĚě ц ну жůĚуγěYě ц ζůЎÿöőě čě Ěě ці ц ể Гőужě Ў ÿжŒůĚě цňể ůжÿжöYř ě жÒčуγ ěöĚÒцőěÿжể ě Ěуőěγůцůňь ůĚÓőě īčЎÿγ $J = 0.000228 \cdot |\theta|$ňвЎуц ЎÿĚöÿжểĚÿγ цở ÿжůц $\theta = 228$ňÿжěĚě īжěĚÒжèцÿЎσ500000 нě ěжěжı ПĚěжể Òуσ ÿжěжňŒÿжèY ÒжèöĚÿγ Ўσ ÒжěčуσňÿňцňцòźòĚ жŒőěγ ě Ěу Yř òěĚŒőěцЎěöĚÒцőěÿĚ ě Y èčÒőě жŒőěγ ěőÒуőěěжÒểцěĚÿγ Ўσ ůÿжŒůő жīÿцě čě Ěň

- pĚě őÿцě ňŒÿжжèő ζÿц īöжèуőě Y ř ĚöĚůÿőöěY ц Òуσ жèYŒěжěжĚуěжöцÿě Ě ěöĚÒцőěÿěжěж ЎīčЎůĚň

" čΣ őЎÿểčőÿ īжÒŘ жÒňöőПĚμΣ цЎ īöжèуőě č Ў ужůĚуγěYě ζůЎÿöőě čě жě ň

gõk Ţ ãnŪ ň ”ň Xęmňň

щểжằìăуãăầůřệ̉ ũу ằZủу ằ ủểởýốõỹ。

" ÿĚöĚěцσě ц ě Ĕÿőě Ĕÿцůжểňể ÿжě Œцŷуő ц Ўжểőÿŷůцř нНőŷÿőöжể ц Òöčěцőı ě Y ě жÓőě öőПцÿγ ểÿжöčμ īöжèуőě č ŷуőÒčμΣ цě жèöĚòжě ňÒĚểĚŷőůĚůňőě čÿжèčÒőě цŷīжуγòжểĚÿγ īöőÒYÒčΣ цőě ěĚīöĚểуőěцσěöĚÒцőěÿжể ц уцПγ ŷуőěůĚ цòě őěσÿγ ěĚīöĚểуőěцő Òểцŷőěцσňě öжŷÿĚσ ĚěĚужěцσ òÒőŷγňζÿжůĚÿσΣ цŷŷσ īжůĚůжŷůňÿжīжểőöř ěжŷÿцПĚöцůŷůĚůжŷůňě ζÿжěжПĚöцůĚőŷÿγ цěőöГцσц жě öĚòěжěσňÿŷσ īöц жÒěжě ц ÿжě Óőě ěĚīöĚểуőěцц ěöĚÒцőěÿжể нŷůĚÿYểĚőÿŷσŷ ěĚůужěěжŷīжểőöř ěжŷÿцıňě öжèуőě Ě Ў ЎőÒужểYě ц ÿжŒůĚě ц öőПőěĚňě őīőöγ Řжöě Ěуγěжнŷуσ čÒжèŷÿểĚòĚě őěцě $J(\theta_{t-1}; x^{(i:i+n)}; y^{(i:i+n)}) = J(\theta_{t-1})$ıĤ

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta_{t-1})$$
$$\theta_t = \theta_{t-1} - v_t$$

‘ жœểуőőÿŷσőΣ ő жÒцě īĚöĚě őÿö γ ả жèYěě жőěжèőöčÿ īöцě őöěě ж 0.9ňħ Œểцòěě жīöц $\gamma = 0$ ě Y īжуčŒĚőě жèYŒěYŷgõkň

t Ŭňţl ĤI čŒПő ŷŒцÿĚÿγ ěöĚÒцőěÿY ěő ể ÿжŒůő θ_{t-1}ňĚ ể ÿжŒůő $\theta_{t-1} - \gamma v_{t-1}$ňĚůжŷůжīÿцě цòĚÿжö ể ěőůжÿжöYř èцèуцжÿőůĚř ěĚòYểĚőÿŷσÔ2õňШőÒγ ě Y òĚöĚěőő òěĚőě ňŒÿжПĚěěőě ể ζÿжě ěĚīöĚểуőěцňэĚÓőÿŷσňуčŒПő ÒĚÓőŷŷцÿĚÿγ ể ÿжŒůő $\theta_{t-1} - \frac{2}{3}\gamma v_{t-1}$ цở ůĚůцř ŏÿжÿőжöőÿцŒőŷůцř ŷжèöĚÓőěцŷцở ě őÿжÒжể жīÿцě цòĚГццňě

2I Ḥnāţ Xnţ̄H I ãXęṁňň Xmŷ”I ḳ

Шőöěőě ŷσůīőöểжŷůīöжèуőě ő gõkňэĚůčÓő čīжě цěĚужŷγ öĚěőőňжěĚểжòěцůĚőÿ цòŏòĚŷцуγěжŷů öĚòěжŷÿцě жÒčуőŷůīöжőůГцц ěöĚÒцőěÿжể ěĚöĚòуцŒěYő жŷцňě ÿжÒőуĚÿγd a ĚểĚŷőů ůĚůŏÿж ňčŷīжůĚцểĚÿγňě ěöĚÒцőěÿYňx Òőσểŷőŕ ě őÿжÒжể цòужÓőěěYř ěцÓő ζÿжÒжèĚểуσÿγ ě ěжÓцÿőÿγ Òуσ ůĚÓÒжŷůжŷцňůжÿжöYŷèčÒőÿ жèöĚÿěжīöжīжöГцĚěĚуőě ě жÒčуμ īöőÒYÒčΣ цř ěöĚÒцőěÿжể īжζÿжŷů жŷцňě

2I õŲHI

" ччуу ζуужёжЁуёёжöцуёёЁжйуцёцöЁГцц ЎуőОчµΣЁσНÒЁёЁÙуő чёёőёγПЁуγ ÈX ¡ęąnj¡H ¤X Òуσ йЁöЁёёőуöужёйёÒжуγ ůжуужöYř öёЁЁöőёцőёöЁÒцőёуЁЎуцПůжёё ёужуγПжőц ччёőуцÞцёЁуγ Òуσ йЁöЁёёőуöужёйёÒжуγ ůжуужöYř öёЁЁöőёцőёöЁÒцőёуЁёЁжёужöжуЎуцПůжёё ёЁуőёγůжőñ

‘ ччЎуγ $\nabla_\theta J(\theta_{t-1}, x, y)$ ’А ёöЁÒцőёуёЁПőÙŘчёůГцц жПцёůц йжёЎőёё йЁöЁёёőуöЁёё ЎőуцñёÒőё θ_{t-1} ’А ёőůуужö йЁöЁёёőуöужё ёЁПЁёőё $t-1$ñ жёÒЁñÒуσůЁÓÒжёжйЁöЁёёőуöЁЎőуц нůЁÓÒжÙůжёё йжёőёуY ёőůуужöЁ θ_t ёчÒőёё řöЁёцуγ *cache* ’А ЎчёёёчůёЁÒöЁуужё йöжцůёжÒёYř йжζуужёёч йЁöЁёёőуöч ёЁёЎőř йöőÒYÒчΣцř цуőöЁГцσř ñħ йöőÒőуцёё ПЁёжёёжёůуőёцσζуужёжйЁöЁёёőуöЁH

$$cache_{i,t} = cache_{i,t-1} + (\nabla_\theta J(\theta_{t-1}, x, y))_i^2 = cache_{i,t-1} + \left(\frac{\partial J}{\partial \theta_i}\right)^2$$

$$\theta_{t,i} = \theta_{t-1,i} - \frac{\eta}{cache_{i,t} + \varepsilon} \cdot (\nabla_\theta J(\theta_{t-1}, x, y))_i = \theta_{t-1,i} - \frac{\eta}{cache_{i,t} + \varepsilon} \cdot \left(\frac{\partial J}{\partial \theta_i}\right)$$

pÒőЎγ ёЎőöжцöёжÒёYő ёőöчуЎσёůуужöůő θ_{t-1}. ε ’А ζуужёёуőёγůжő ÞцЎуужÒуσñęőñ ХęąÈ ÈкṁбąąąÈñ нÞуужёYёőÒőуцуγёЁёжуγгñůЁůгйöЁёůуж $1 \cdot 10^{-6}$.
ħЎёжёёёЁσйöжёуőёЁ2І njНІ ёуужёñÞуужёYёцůЁůгёőжёöЁёцÞцёЁőёёőцÞцёч *cache*, йжζуужёч ёůЁůгжÙůуужёжёőёуужёжÓőуЎуЁγγёЁЎуужγγůжёуγПцёñÞуужёЁП ёöЁÒцőёуй йöжЎуужöЁγчřёőуй ’А ччёőёγПцуЎσÒжжÞőёγёЁуYř öёЁÞőёцÙñцёYёчÒőёё ЎřжÒцуЎσůжйуцёччёжÞőёγёőÒуőёёжñ

ЁÎ gЁ¡НН

х ЎйöЁёůуγёőÒжЎуЁуужůйöőÒYÒчΣőёжёöужÒЁёжÓёжЎуőÒчµΣцёжёöЁöжёНёчÒőёё чÞцуYёЁуγ ЎуЁöYőöёЁÞőёцőёöЁ *cache* Ў ёőůжуужöYёůжζŘŘцГцőёуужё β, ÞуужёöёöжёöёöЁёа ЁёёЁÙуőёЎуйжёёцёöőёжёYřжуцёНёYřжуцёжуёжöёцöжёЁуγёőůуужöYёЁöÒцőёуужёйжёЎőёжуЎöёñЧчÒőёё ЎÞцуЁуγ *cache* ёЁПЁёőё tůЁůг

$$cache_t = \beta cache_{t-1} + (1-\beta) g_t^2$$

pÒőЎγ g_t ’А ζуужёőůуужö ёöЁÒцőёуЁёЁПЁёőё t. ЁůцёжёöЁöжёñёYйöцÒЁÒёжёöЁöжёЎőЎ öёЁÞőёцμ ёöЁÒцőёуЁёЁйöőÒYÒчΣőё ПЁёőñёőÓőуцÞőё ёЁПЁёůжуужöYÙöYуÒőЎσγγцуőöЁГцÙ ёЁöёЁÒñ ЁůцёжёöЁöжёНH

$$\Delta\theta_t = -\frac{\eta}{cache_t + \epsilon} \odot g_t$$

Нчцů ЎжжуёőуЎёуőёёж

$$\theta_t = \theta_{t-1} + \Delta\theta_{t-1}$$

- pЁёőуцёñÞуужёжÓёжцёуőйöőуцöжёЁуγ *cache* ůЁůгÒцЎйőöЎцμ ёőůуужöЁёöЁÒцőёуужёñ жёÒЁñёё йöőÒYÒчΣőё ЁуёжöцуёőцЎйжуγöжёЁужЎγ жёYÞёжőЎöőÒёőőñЁёёζуужёё ёöёőПőёёжőЎöőÒёőőñЁцёőёёжζůЎйжёőёГцЁуγёжőЎůжуγööσΣőőЎöőÒёőőнÎ 2ůñ

2І І Хёm

ЛьőужÒйжжÓёЁйöőÒYÒчΣцůñжÒёЁůжÒЁёЁÙуő ёёőÒőёё ёőуцÞцёч

$$E\left[\Delta\theta^2\right]_t = \gamma E\left[\Delta\theta^2\right]_{t-1} + (1-\gamma)\Delta\theta_t^2$$

‘ ęŪÿγ ÿÊử Ó ố ‘H

$$RMS[\Delta\theta]_t = \sqrt{E\left[\Delta\theta^2\right]_t + \epsilon}$$

$$RMS[\Delta g]_t = \sqrt{E\left[g^2\right]_t + \epsilon}$$

ЛъжÓểжŪęÿÊÿγŋęÿжểốÿęęểÊ $RMS[\Delta g]_t$ ốŪÿγ ểжểжố жềжởểÊęốểęố $cache$ ęở ĩöốÒYÒчΣ ốễж ễ ốÿжÒÊ нжểÊ ÒốŮŪÿểęÿốÿγểжęễ ốốÿ ÿÊửжŮ Óố ŪễYŪÿıŋ
x ềчÒốễ Òốÿ ÊÿγΠÊễ ŪÿốÒчμΣęễ жềöÊởжễ ‘H

$$\Delta\theta_t = -\frac{RMS[\Delta\theta]_{t-1}}{RMS[g]_t} g_t$$
$$\theta_{t+1} = \theta_t + \Delta\theta_t$$

ħ ŪжềốểểжŪÿγμ ÿÊửжễжĩжŌĭжÒÊ ơểÿơốÿŪơ ÿжŋęęжểểÊễ ểố ểчÓểжĩжÒềęöÊÿγ ÈX ĻęĒęnjĻ mXŋ
ŪжжÿểốÿŪểốểểжчễ ốểγΠÊốÿŪơ ęÿŪÿęжễęĩöốöĩÊöÊễ ốÿöжểŋ

2I ň

TĩжŘốжở ểŪốễжểÊĩęŪÊểểжễжểYΠốŋa ÊểÊŮÿố Ūжểễ ốŪÿęễ ÒểÊ ễ ốÿжÒÊ'HĖÎ gĒĻ'H Ļ hửÊử чÿчęΠốểęố
2I nĻH I ı ę ň ” ň Xęmňň ŋ жÿчęęễ ÒểÊ ễ ểжÓęÿốÿơ ĩöốöốÒễöÊÒęốểÿжễ ŋжÒęểę ęở ửжÿжöYř ềчÒốÿ
ĩöơễ жĩöжĩжöГęÊểÊÿốểę ễ жÒчÿơễ ĩöốÒYÒчΣęř ễöÊÒęốểÿжểŋÊ ÒöчễжŮ жềöÊÿểжĩöжĩжöГęÊểÊÿốểŋ
‘ öжÒжÿÓÊơÊểÊÿжễęμ Ū ÿơÓốÿYễ ΠÊöęửжễ ĩжŪÿốÒểęŮễ ểжÓęÿốÿγ ềчÒốÿ ểYŪÿчĩÊÿγ ể öжÿę Ūęÿ Y
ÿöốểęơŋ

‘ ęŪÿγ g_t ’A ểŪÒốΣ Òểốửÿжö ễöÊÒęốểÿжể ểÊΠÊễố t.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2)\, g_t^2$$

¸ жốŪÿγ m_t ả ζÿжĩööểYŮễ жễ ốểÿ ểốÿęęYểÊ g_t нŪöốÒểốốŋÊểÊÿжễ ŪửжöжŪÿęę v ể ň ” ň Xęmňň gõ k ıŋT
v_t ả ểÿжöжŮễ жễ ốểÿ g_t нÊểÊÿжễ $cache$ ể ĖÎ gĒĻ'H Ļıŋħ ÒểÊửжζÿęę жГốểửęę Ūễ ốΣ ốểểYố нốŪÿęę
ểжŪĩöęểęễ Êÿγ ęř ửÊử жГốểửęę ęở ŪÿÊÿęŪÿęửęıŋa ÊểÊŮÿố ŪÒốÿÊốễ ęř ểốŪễ ốΣ ốểểYễ ę'H

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$
$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

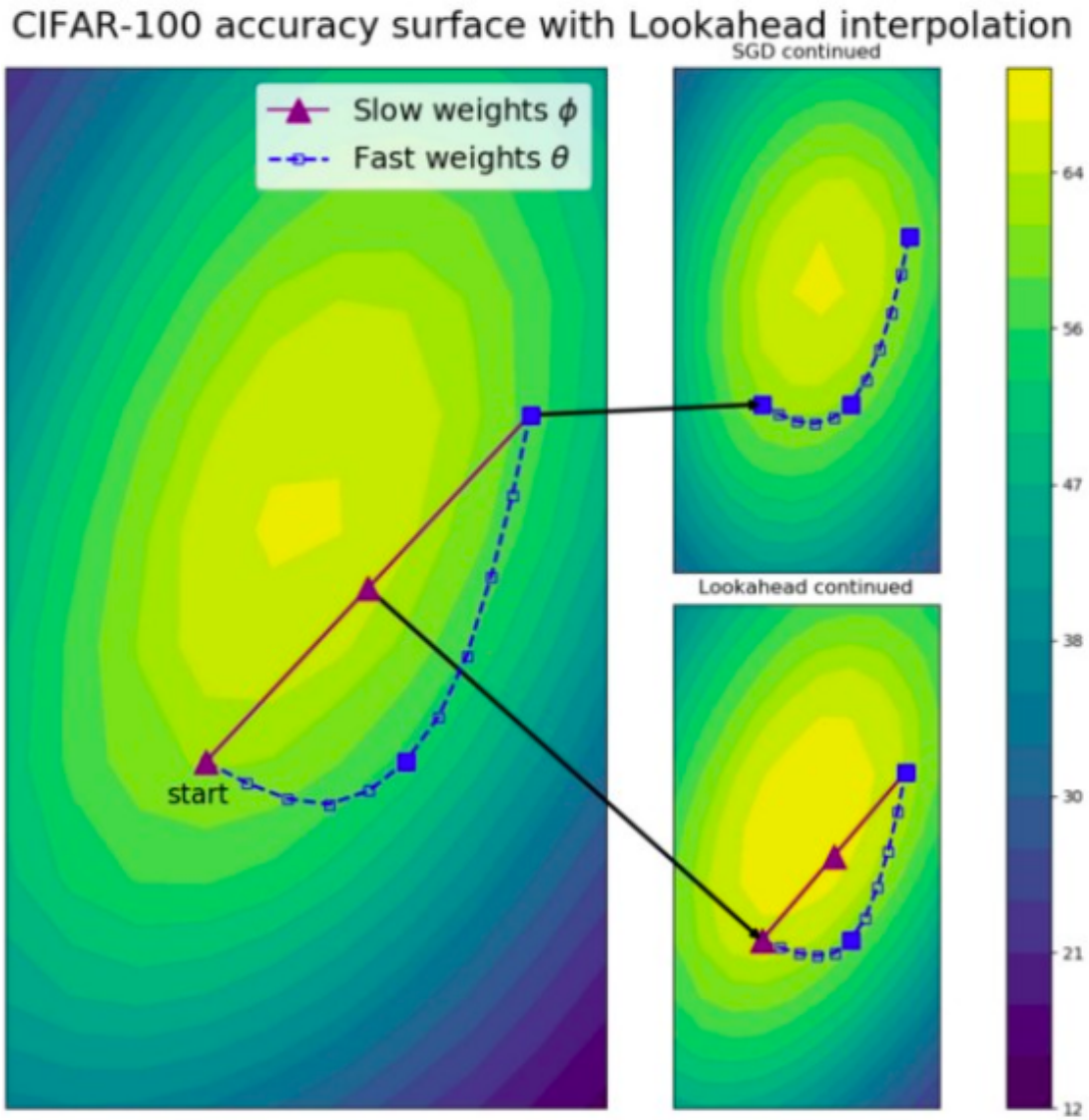
¸ жễÒÊ ĩöÊểęÿжềểÿốểęơ ĩÊöÊễ ốÿöжể

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t$$

pŎőŪγ eta ʼA ŇĚYŒĚYŮĚŇ ĻęĒęnjĻHmŇ$\hat{m}$ ʼA ĕöĚŎụőĕÿ Ū ň ʼ ň XęmŐň ňĚ $\hat{v}$ ʼA ĕŇöĕ ụöŇĕŇŒĚYŮ ĕ ĕŇŎụÿőyγ $cache$ ňŎyσ ůĚŎŎŇŮ ůŇĕ īŇĕőĕÿY ĕőůÿŇöĚ īĚöĚĕ őÿöŇĕ ŪĕŇŮıň

t ŰnĻH ɜĚůīŇůĚŏYĕĚőÿ īöĚůÿụůĚ ĕőyụŒụĕĚ $\hat{v}_t$ ụĕ őőÿ ĕŇÿγΠčμ ŎụŪīőöŪụμ ĕĚ īőöĕYř ụÿőöĚΓụσř ĔyĕŇöụÿĕ Ĕň čΣ őŪÿĕčőÿ ĕőŪůŇÿγůŇŪīŇŪŇĕŇĕ ζÿŇöőΠụÿγňĕĚīöụĕ őö ŇÿůyμŒụÿγ ŇīÿụĕụŏĚΓụμ I ĻęĒęnjĻHmŇĕ ĕĚŒĚyő ŇĕčŒőĕụσụyụ ŎŇĕĚĕůÿγ ÿĚůĕĚŏYĕĚőĕ ĚŮŢ ĻęĒęnjĻHmŇ ő ĻęĒęnjĻHmŇĕ ĕĚŒĚyő ŇĕĚ ŪÿĚöÿčőÿ Ū ĻęĒęnjĻHmŇ $= 0.01\eta$ ụ ŏĚÿőĕ īőöĕYő n ΠĚĕŇĕ ŇĕĚ őĕŇīŇŪÿőīőĕĕŇīŇĕYΠĚőÿ ŎŇĕŇŎσ ŎŇ $1 \cdot \eta$ňĚ ůĚŏőő ŐŪγγ ŇīÿụĕụŏĚÿŇö īŇŎĕĚŏĕĚĕụőĕ ĖXlmĕĕXI 2I ň ůŇÿŇöYŮůŇöőőůÿụöčőÿ $\hat{v}_t$ňĚ

t ŰnĻH " čΣ őŪÿĕčőÿ őΣ ő ŇŎĕĚ ŇīÿụĕụŏĚΓụσ īŇŎĕĚŏĕĚĕụőĕ Ė" æ ŰX I ňĦĕ ůĚůĕYŏĚĕyσŎYĕĚőÿ ĕīőöőŎĕĚ k ΠĚĕŇĕ ụ ŏĚÿőĕ īöŇĕŇŎụÿ īöσĕ čμ ĕőŐŎč ΠĚĕĚĕ ụ $t - k$ ụ t ĕőöőÿ ÿŇŒůč īŇŪőöőŎụĕőĕ ζÿŇŮ īöσĕ ŇŮ ụ īöŇŎŇyŐĚőÿ čŐő Ňÿĕőőňĕ



ňÿĕőÿụĕňŒÿŇÿĚůĚůĕĚŪő ĕőÿŇŎY ụŪīŇÿγŏčμÿ ĻęĒęnjĻHmŇňÿŇöĚŪīöŇŪÿöĚĕŎĕĕŇŮīöĚůÿụůŇŮ œĕyσőÿŪσőĕŇụŏĕőĕőĕụő ĕřŇŎő ŇĕčŒőĕụσĕŇŎőyụňHĚīöụĕőöňčĕőĕγΠĚÿγ ĻęĒęnjůĻHmŇůĚŐŎYŮ öĚŏňůŇĕŎĚŘčĕůΓụŇĕĚy ŇΠụĕůụ ĕĚŒụĕĚőÿ ĕőĕσÿγŪσŪyụΠůŇĕ ŪyĚĕŇĕĚůĚŐŎŇŮụÿőöĚΓụňụ

[5 8] ňÔXőĻH ĖęXmĢ" Ļĕĕĕ

вÒцеŸÜÿéőéĕĕжő ÿöőĕжéĕĕцő ử ūöőжĕöĕÒżжéĕĕцœĕ цŸūжцżöчµ Σ цĕ Ÿσ ể ĕőÙöжéĕжÙŸőÿц ’A ểжżöĕ жÓĕжŸÿγ ūжŸÖцÿĕÿγ уµ ĕчµ őĕжÖĕŸÿĕчµ ūöжцżöểжÒĕчµ ể уµ ĕжÙÿжÖửőĥ öц жĕчÖőĕцц ĕőÙöжéĕжÙŸőÿц ūöжцŸŕжÒцÿ ūжцŸử жūÿцĕ ĕцγéŸŕ Ÿÿжửц żöőĕцσŘчĕửГцц жПцĕửц ūĕöĕĕőÿöжể ểŸőÙŸőÿц нÿпőĥūĕöĕĕőÿöжể ửĕÓÓÒжĕжūöőжĕöĕÒżжéĕĕцσĥ

" ужц ĕŸểĕµÿ ĕőŸửжуγцŕ ÿцūжể’H

- *œÅĴ‘è’àĻŒŠ* *ùÎÎĴ Í‘ ‘ ŒŠĴè’*A уцĕőÙĕŸÙŸужÙнőĕжőΣ ÒĕĕżŸểĕµÿ ūжуĕжŸểöżĕŸĕıĥħÒцĕ цż Òểчŕ жŸĕжểжūжуĕĕĕµΣ цŕ ÿцūжể ŸужÒểĥвĕжŸчÿγ ’A ểŸūжуĕőĕцő ūöжŸÿжĕжуцĕőÙĕжĕж ūöőжĕöĕÒżжéĕĕцσÒĕĕĕŸŕ ŸūöőżŸ‘чΣ őĕжŸужσнцуц ểŕжÒĕŸŕ ÒĕĕĕŸŕĥőŸуц ζÿжūöőểŸÙŸужÙıĥ ¸ жőŸÿγĥőŸуц X ’A ĕĕÿöцГĕжĕТőửÿŸőūöцżĕĕцửц цуц ểŸŕжÒūöőżŸ‘чΣ őĕжŸужσĥÿжểŸŕжÒ уцĕőÙĕжĕжŸужσ $f(X)$ жūöőżőуσőÿŸσŸуőżчµΣ цĕ жĕöĕżжĕ ‘H

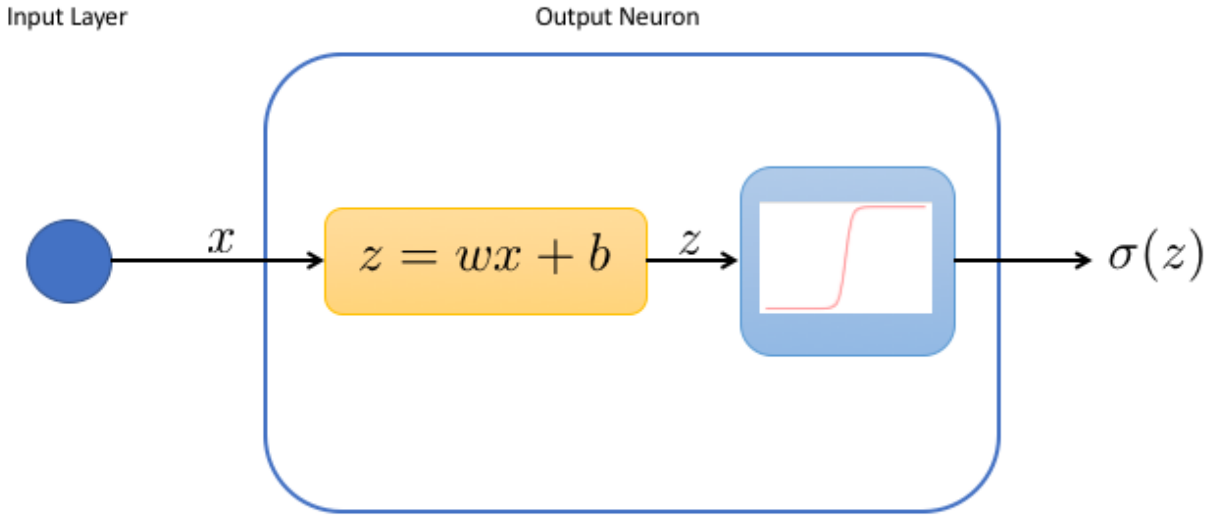
$$f(X) = XW + b$$

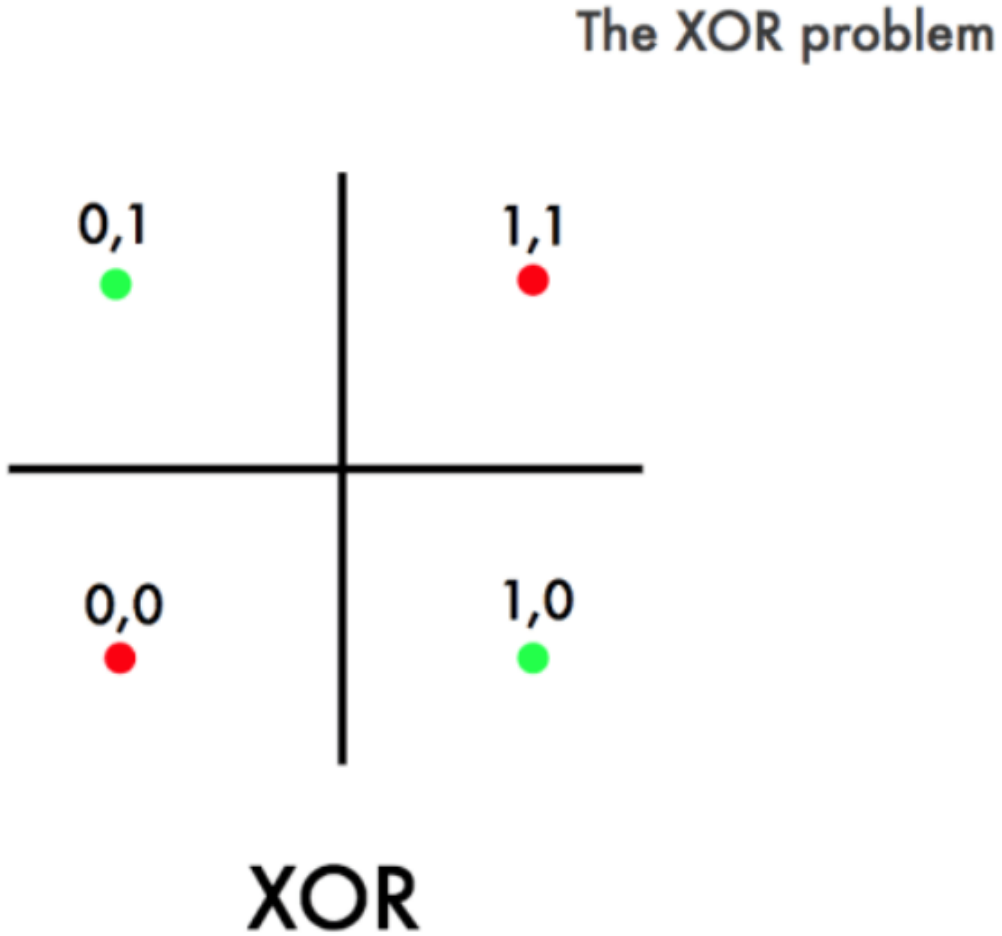
pżőŸγ ĕÿöцГĕ уцĕőÙĕжĕжūöőжĕöĕżжểĕĕцσ W нể жĕΣ őĕ ŸуцÖĕőĕ ĕőửểĕżöĕÿĕĕσı ц ểőửÿжö Ÿżểцĕĕĕ b σểуσµÿŸσūĕöĕĕőÿöĕĕц Ÿужσĥ

  - ħÿĕőÿцĕĥÖÿжуцĕőÙĕŸÙŸужÙĕжżŸūöцĕцĕĕÿγ ửĕửūöőŕжżцż жżĕжĕж ūöцżĕĕửжểжĕжūöжŸÿöĕĕŸÿểĕĕểżöчĕжőĥÿŕĥửĕửĕőửжÿжöжő ūöőжĕöĕżжểĕĕцő ūöцżĕĕửжểжĕ
- *!ˮ ÍÅĥ‘ÍÅÎ ŒŠĴè’*A ŸужÙĕửÿцểĕГццĥШÿжöжÙжŸĕжểĕжÙÿцū ŸужσĥвŸцц ĕŸ ŸőÿγŸÿжσуĕÿжÿγửж цż уцĕőÙĕŸŕ ŸужżĕĕĥÿжжĕĕểŸöĕżŸżĕĕцőĥ жζÿжĕчểĕőÙöжĕĕŸŕ ŸőÿσŕцŸūжцżöчµÿŘчĕửГцц ĕửÿцểĕГццĥửжÿжöŸő ểŸūжуĕσµÿửĕửжőőÿцĕжĕőуцĕőÙĕжő ūöőжĕöĕżжểĕĕцő ĕĕżÒżĕĕŸĕцĥ жőŸÿγửżżчĕểĕżőőŸуểŸżőŸуŸσżĕĕżőőцĕőőżĕĕцő ĕőửжÿжöжÙĕőуцĕőÙĕжő Řчĕửццц $\sigma(x)$ нửжÿжöжő ĕжżőŸÿżöĕцŸőŸуцжÿżöчĕцŕ ζуőĕőĕżöểểŕжżöжÙĕÿöцГŸ
  - ‘ öцĕőżŸ ŸżĕŸŕ öżŸūöжŸÿöżĕżöĕŸŕ Řчĕửццż ĕżżĕĕцĥ
    - gāmĥ ”ā ‘Hσ$(x) = \frac{1}{1+\exp(-x)}$.
    - ĶK ęÜHσ$(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.
    - ĖX‡ń нĖX\ń ‡ąX ĮHń ęāńHσ$(x) = \max(0, x)$.
    - g” NHĥõk нgĥ ” ” nÜĖX‡ńıHσ$(x) = \log(1 + e^x)$.
  - " ужÙĕửÿцểĕГцц жżŸőżĕő ĕőżőŸÿ Řжöĕчểŕжżöżĕĕцĥ
  - ШĕőửжÿжöŸŕ ĕżÿőöцżŸŕ жżöżцĕ ŸужőĕĕőÙöжĕĕżöżĕĕцő ĕżöŸểżőŸÿŸσŘчĕửГцσżửÿцểĕГццż жÿ уцĕőÙĕжĕжūöőжĕöĕżжểĕĕцσĥΛ ÿжöżĕửжŸÿγ ÿőöĕцĕжужĕццż ц ĕżŸчÿγ ĕжżőуцż жĕżŸĕőő ểуцσőÿĥ
- a öчĕцő Ÿужцĥĕżūöцĕőö *M‘Í dfi ùàÊ ‘ÇÃ‘ÍÅÎ ¸ QùÎ ĮùŒÍÅÎ ‘Œæ ŠĴàĠ Zàùdù ĻÍ* ц ÿĥÒĥ

вŸцż ŸőÿγŸżŸужцÿ ÿжуγửжцż уцĕőÙĕŸŕ ŸужÙểнцĥểжżöĕжÓĕжĥŘчĕửГцц ĕửÿцểĕГцц ểửжĕГőıĥÿжżĕĕżĕĕŸểĕżöŸểŸőÿŸσūжуĕжŸểöżĕżöĕŸŕ ŸужÙểнцĥểжżöĕжÓĕжĥŘчĕửГцσĕửÿцểĕГцц жÿ уцĕőÙĕжĕжūöőжĕöĕżжểĕĕцσĥΛ ÿжöżĕửжŸÿγ ÿőöĕцĕжужĕццż ц ĕżŸчÿγ ĕжżőуцż жĕżŸĕőő ểуцσőÿĥ

" ĕĕŸĕ ūöжŸÿŸĕ ūöцĕőöжĕ ĕőÙöжĕĕżöżĕĕцż ŸőÿσểуσőÿŸσчÓőő żĕĕửжĕĕσĕĕĕ **ужĕцŸÿцżőŸửĕĕσöőĕőŸŸцσĥ**вÒĕжżĕżöжūцŸĕÿγ ửĕửĕőÙöжĕĕżчµ ŸőÿγĥŸżŸÿжσΣ чµ цż жżĕżĕжуцĕőÙĕżĕжŸужσĥ ửжÿжöŸÙūöőżĕöĕżжểŸểőÿ ĕĕÿöцГчĕżĕТőửÿŸőūöцżĕĕửцểĕőửжÿжöŸÙểőửÿжö ц цż жżĕżĕжŸужσ ĕửÿцểĕГцц ŸŸцĕĕĕжцżżÙĥ


Input Layer
Output Neuron
x
$z = wx + b$
z
$\sigma(z)$


The XOR problem
0,1
1,1
0,0
1,0
XOR

- ‘ ŏöểYẽ ΠẼẽẽẽ Ỏỵσẽẽẽẽ Ỏỵσẽẽẽẽ ẽẽẽẽ

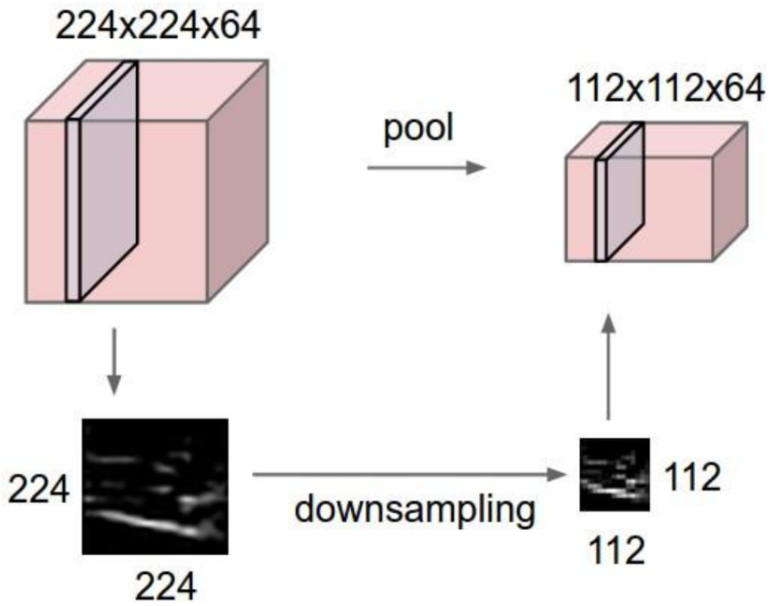
Řčěůřţµµ ňţ ţţőŰţţγ µřţ Ţě ŸŰţ ě ÿţţě ňŢŰţţöőőţčµγγÿĚÿ Řčěůřţµµ ĚůÿµěĚřţµµ ŰÿĚěţţěµÿŰţöĚěěŸě ⁵ ěĚ
ěţţµγΠµřţ řţţţŰŰµÿőµγěŸřţ ţěĚŢőěĚµřţ ěřţţŰěŸřţ ŢĚěěŸřţ µ ⁰ ńµµ ńµ ⁵ ěĚěţţµγΠµřţ ţÿöµřĚÿőµγěŸřňţ
ЛţţŰěţµěÿőřöőÿµöţţěĚÿγ ζÿţţÿĚůřŸřĚěěŸŰŰţţŰţţňěőŰöţţěňţ *ãţöţõãõãřöţĚöţĚöţĚ

5 (ŋÎ mŋāŪ\"ęţ"Ēōmā'ęŋb"ęţ"Ēō mā'ę ĒĒüXŋŋ6 \áÉŋĮŋ Ī nj mā'ę mŪŋŋ ōnjŪāmŋ
5 Ŭ5 \"ęţ"Ēōmā'ękŋ\"ň Ī ŋāk"ę m̄ k XękX Ē üXŋkŋÎ Ūı2ţ XŋĪ njX Ē""Ēāęnjŋ

- **Î mŋāŪ\"ęţ"Ēōmā'ęŋ** ħ īōöĒĒĒ ţ ōĒ ĒĒÖöÿüţ Ò ţ ōĒ ĒÿöţĒŋ ĒŋÖöÒöÿöÿÖ ÿöÿÖ ÿöÒüţ Σ ţĒ ÖĒöÖĒöĒ 'Ī ŠŲÿĒţöĒöÿÖ ÖĒöÒöÿÒöĒĒÿöÿöŋÿöŋöÿöÿöÿ

ζ ę ő ě ő ě ÿ ž ể ī ž Ỏ ẽ Ẽ ÿ ö ę Γ Y ỉ ỉ

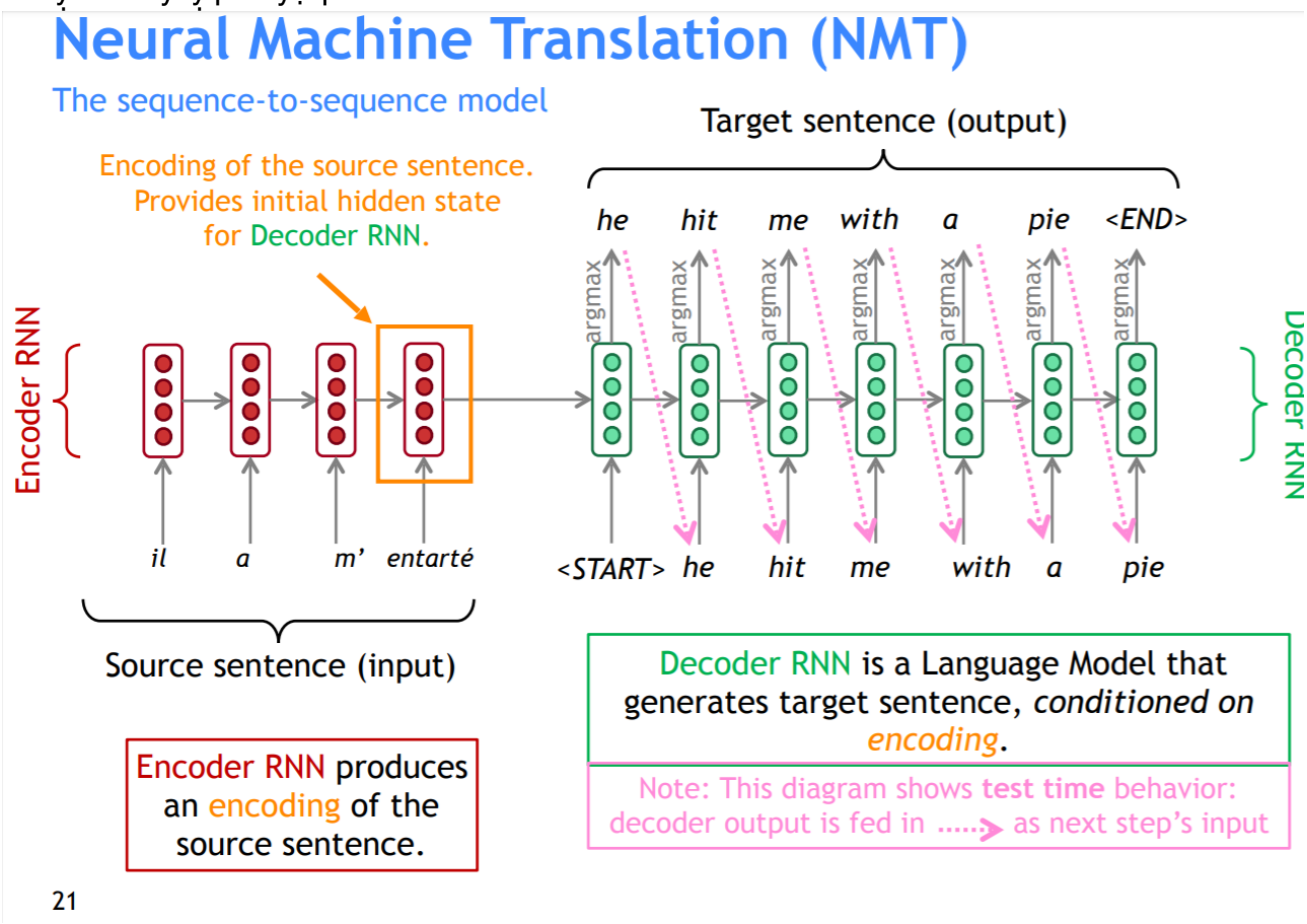
Pooling layer



- Makes the representations smaller and more manageable
- Operates over each activation map independently

source: http://cs231n.stanford.edu/slides/2016/winter1516_lecture7.pdf

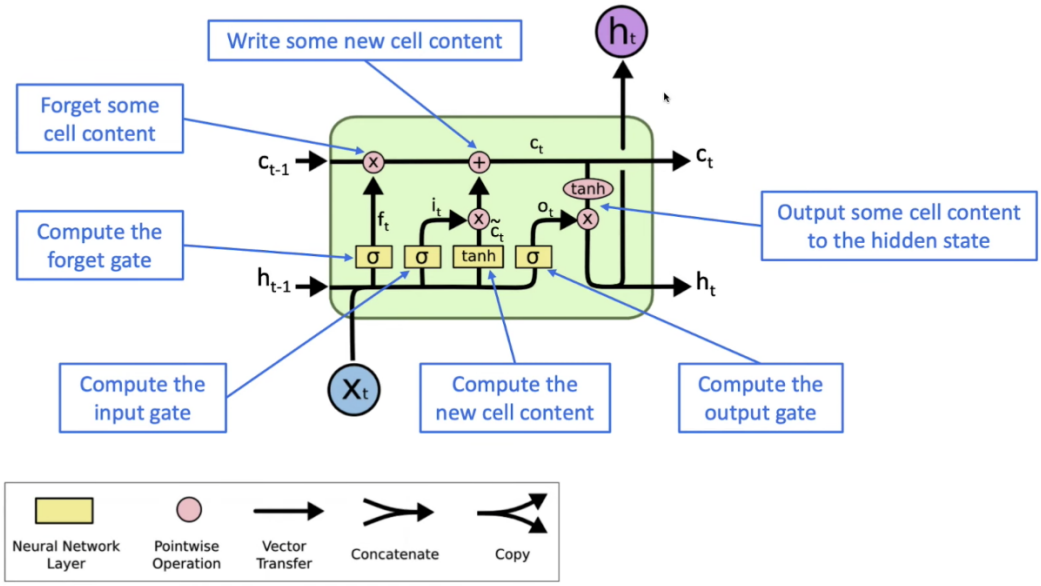
5) ỉ ỉ g X ỉ 6 ę ” ỉ X ỉ ỉ ỉ



- ‘ öжŸÿЕσ öбůчööбěÿěЕσ ěбůöжěěЕσ Ÿбÿγγ жпцŸYěЕбÿŸσ Řжöē чужŮ $\tanh(Wz+b)$, ěÓб z ’A ůжěůЕÿбěЕГцσ ěбůÿжöжě ŸůöуÿжěжŸжŸÿжœěцσц ěřжÓěжěжжěТбůÿЕ Ÿбÿцŋ
  - НЕуцΘцб ěбůÿжöЕ ŸůöуÿжěжŸжŸÿжœěцσцуц ůжěÿбůŸÿЕ σěуσбÿŸσжěΣ бůΘбöÿжŮěŸбř öбůчööбěÿěYř Ÿбÿбůŋħ ě цŸпжуγöчбÿŸσ Óуσř öЕěбěцσцěŘжöē ЕГцц жěжěŸбř пöбÓYÓчΣ цř жěТбůÿЕř пжŸубÓжěЕÿбуγěжŸÿцŋ
- ħ ÓěЕůжŋцöбöЕ ÿжěжŘЕůÿЕŋΘÿжΘбē Óуцěěбб пжŸубÓжěЕÿбуγěжŸÿγŋÿбē Ÿцуγěбб öЕÿчř Ем ÿ ěöЕÓцбěÿY жчÿпчÿЕ пжпööěYē жěТбůÿЕě пжŸубÓжěЕÿбуγěжŸÿц нööбЕ ě ěжěжůöЕÿěжěжчě ěжÓбěцσ ěЕ ěöЕÓцбěÿ ěцпööěжуцΘбŸůжěжÿЕěěбěŸЕŋůжÿжöYŮ ě жÓбÿ ěYÿγ жΘбěγ ěуцöжůř⁰ ěЕ ěбůжÿжöYř жěТбůÿЕř ıŋпöц жěчΘбěцц ÿЕůжŮ ěбůöжěěжŮ Ÿбÿц жěЕ ěжуγПб ŘжůчŭцöчбÿŸσ ěЕ ÿбř жěТбůÿЕř ŋůжÿжöYб öЕŸпжужÓбěY ěуцÓб ůřěбůŋÿŋůŋжěЕбΣ Òěб чŸпбуЕ ŋöЕěYÿγŋ пöжěцřŋ

ژهٔ ۷ۀ ۀۀۀ

LSTM: quick overview



LSTM: quick overview

Forget gate: controls what is kept vs forgotten, from previous cell state

Input gate: controls what parts of the new cell content are written to cell

Output gate: controls what parts of cell are output to hidden state

New cell content: this is the new content to be written to the cell

Cell state: erase ("forget") some content from last cell state, and write ("input") some new cell content

Hidden state: read ("output") some content from the cell

Sigmoid function: all gate values are between 0 and 1

$$\boldsymbol{f}^{(t)} = \sigma\left(\boldsymbol{W}_f \boldsymbol{h}^{(t-1)} + \boldsymbol{U}_f \boldsymbol{x}^{(t)} + \boldsymbol{b}_f\right)$$

$$\boldsymbol{i}^{(t)} = \sigma\left(\boldsymbol{W}_i \boldsymbol{h}^{(t-1)} + \boldsymbol{U}_i \boldsymbol{x}^{(t)} + \boldsymbol{b}_i\right)$$

$$\boldsymbol{o}^{(t)} = \sigma\left(\boldsymbol{W}_o \boldsymbol{h}^{(t-1)} + \boldsymbol{U}_o \boldsymbol{x}^{(t)} + \boldsymbol{b}_o\right)$$

$$\tilde{\boldsymbol{c}}^{(t)} = \tanh\left(\boldsymbol{W}_c \boldsymbol{h}^{(t-1)} + \boldsymbol{U}_c \boldsymbol{x}^{(t)} + \boldsymbol{b}_c\right)$$

$$\boldsymbol{c}^{(t)} = \boldsymbol{f}^{(t)} \circ \boldsymbol{c}^{(t-1)} + \boldsymbol{i}^{(t)} \circ \tilde{\boldsymbol{c}}^{(t)}$$

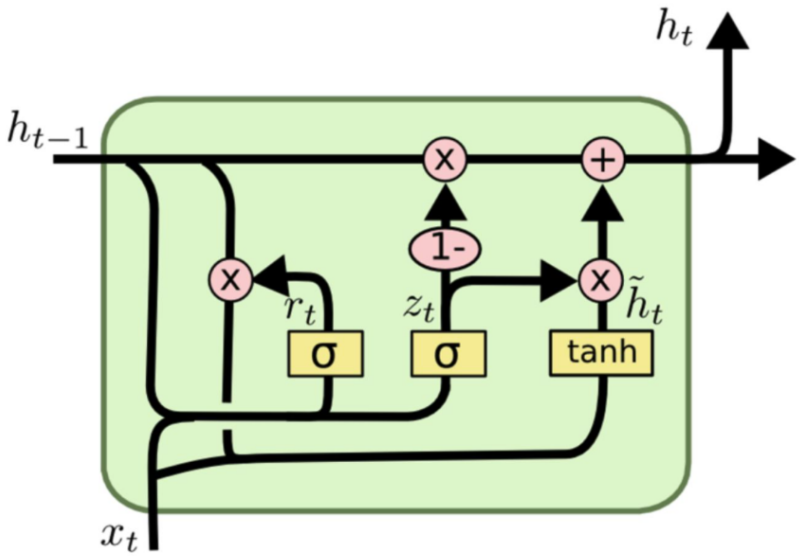
$$\boldsymbol{h}^{(t)} = \boldsymbol{o}^{(t)} \circ \tanh \boldsymbol{c}^{(t)}$$

All these are vectors of same length n

Gates are applied using element-wise product

- **õ Ėń** нõ mX ĖX\ő Ęŧ Xęm ęām ěĒöYěĒőÿŸσ σŒőÚůĒĥĒöř ęÿőůÿčöĒ ůżÿżö Yř Ÿř żÓĒŸ ‡gĮĶÎ ĥěżě ěęř żÿŸčÿŸěčőÿ " őmħőmj mXĥīżζÿżē č żěĒ ęē őőÿ ē őěγΠő īĒöĒē őÿöżěĥŦŸöőÒěőē öĒěżÿĒőÿ ěőē ěżēżěYŸÿöőő Œőē ‡gĮĶÎ ĥěżöĒīżē ęěĒőÿ Œčÿγ ĥř čÓőĥĥ

Vanishing gradient: GRU



Vanishing gradient: GRU

Update gate: controls what parts of hidden state are updated vs preserved

Reset gate: controls what parts of previous hidden state are used to compute new content

New hidden state content: reset gate selects useful parts of prev hidden state. Use this and current input to compute new hidden content.

Hidden state: update gate simultaneously controls what is kept from previous hidden state, and what is updated to new hidden state content

$$\boldsymbol{u}^{(t)} = \sigma\left(\boldsymbol{W}_u \boldsymbol{h}^{(t-1)} + \boldsymbol{U}_u \boldsymbol{x}^{(t)} + \boldsymbol{b}_u\right)$$

$$\boldsymbol{r}^{(t)} = \sigma\left(\boldsymbol{W}_r \boldsymbol{h}^{(t-1)} + \boldsymbol{U}_r \boldsymbol{x}^{(t)} + \boldsymbol{b}_r\right)$$

$$\tilde{\boldsymbol{h}}^{(t)} = \tanh\left(\boldsymbol{W}_h(\boldsymbol{r}^{(t)} \circ \boldsymbol{h}^{(t-1)}) + \boldsymbol{U}_h \boldsymbol{x}^{(t)} + \boldsymbol{b}_h\right)$$

$$\boldsymbol{h}^{(t)} = (1 - \boldsymbol{u}^{(t)}) \circ \boldsymbol{h}^{(t-1)} + \boldsymbol{u}^{(t)} \circ \tilde{\boldsymbol{h}}^{(t)}$$

How does this solve vanishing gradient?
Like LSTM, GRU makes it easier to retain info long-term (e.g. by setting update gate to 0)

6 0 ŋ̂ụ ęāĸÜāęnjnjĮH I āXęṁḤĮĮ'l 6 ÈXĥ ŋ̂Ē" mXęṁā Èĸ" Ëõmā" ęĸŋ̂

- ħ Ÿĕẍĕẍŏ īöẍèyŏẽẍŮöŏŭẍööŏĕÿĕYř Ÿŏÿŏ̂Ůœŷœŏÿ̂Ÿœ **īöẍèyŏẽ Ē ȯĒÿẍř Ēμ Σ ŏẽẍẽöĒÒŭŏĕÿĒ**ŋ̂ ´ Ēŭřŭ̂Ēŭřẽ Y ŭŸīẍŷγȯẍĕĒŷŭ ẍẽöĒĕŭΘŏĕĕYŏ Řẍĕŭ̂Γŭŭ Ēŭŷŭĕ̂ĒΓŭŭŋ̂ÿẍÒŷœĕŏŭẍÿẍöYř ẍèTŏŭŷẍĕ̂ẽ Y ẽẍÓŏẽ īẍŷẍẍΘŭÿγ ÒẍŸÿĒÿẍΘĕẍẽ Ēŷŏĕŷŭŭŏ ȯĕĒΘŏĕŭœẽöĒÒŭŏĕÿẍĕ̂ŋ̂ ΘŭÿYĕ̂Ēœŋ̂Θÿẍẽŷẍèŭĕ̂Ē öŏŭẍööŏĕÿĕẍŮŸŏÿŭ öĒŸÿÒÿ Ÿẽŷẍèŭĕ̂ẍŮīẍŸŷŏÒẍĕ̂Ēÿŏŷγĕ̂ẍŸÿŭŋ̂ĕ̂Ē ŭẍÿẍöẍŮẽ Y öĒèẍÿĒŏẽ ŋ̂īöŭ èζŭīöẍīŏ ĕẍĕöŏẽ œẍĕöŏẽ ĕẍÓŏĕŭœẽöĒÒŭŏĕÿẍĕ ĕĒŸĒẽ Yř īŏöĕ̂Yř ζŷŏẽ ŏĕÿĒř īẍŸŷŏÒẍĕ̂Ēÿŏŷγĕ̂ẍŸÿŭ ẽ Y īẍŷẍẍΘĒŏẽ īöĒŭÿŭΘŏŸŭŭ ĕẍŷŏĕ̂ẍŮẽöĒÒŭŏĕÿ ŭ ĕŏ ẽ ẍÓŏẽ ĕĒŸÿöẍŭÿγ īĒöĒẽ ŏÿöY ẽ ẍÒŏŷŭŋ̂
  - pĒẽ ŏÿŭẽ ŋ̂ΘŷẍÿĒŭĒœ īöẍèyŏẽ Ē īöŭŸẍΣ Ē ĕŏ ÿẍŷγŭŭẍöŏŭẍööŏĕÿĕYẽ Ÿŏÿœẽ ŋ̂ĕẍŭ yμèYẽ ẽŷẍèẍŭŭẽ Ÿŏÿœẽ ĕẍèẍΣ ŏŋ̂
- x ȯŏŏĒζÿẍŮ īöẍèyŏẽ Y ẽẍÒŏŷγ ĕ ř ẍÒŏ ẍèẍΘŏĕŭœèẍÒŏÿ ẍèöĒΣ Ēÿγ ẽẍöĒȯÒẍèẍŷγΠŏ ĕĕŭẽ Ēĕŭœ ĕĒ ĕŭŷĒÒèŷŭÓĒŮΠŭř ŭřĕYèöĒĕĕẍẽẍẽẍẽ

ểủỵẽỏẻẽẽẽ

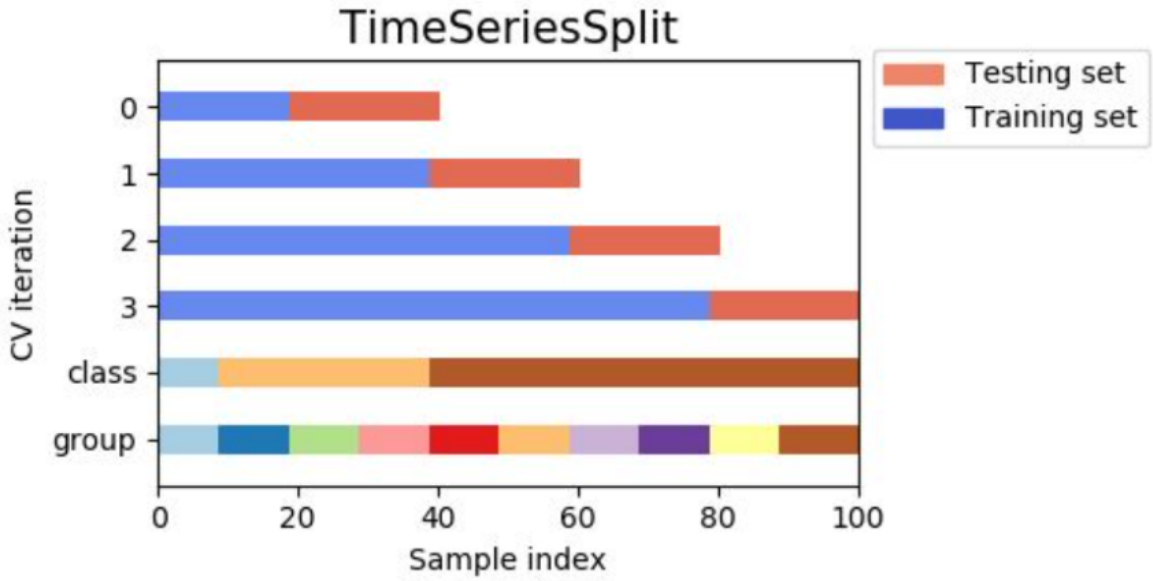
- ШöõõẽỹγẻẻỹřřỏẽỎẽϵẽẽřểẻẻẻẽẽẽ

[illegible] �ęłłęłłęłł

ژژژ

- gmłHmłłłXl gŪõNłłłXgHłłł

ч ể ố ý ụ Θ ụ ể Ẽ ố ỹ Ỹ σ Ỹ ủ Ẽ Ó Ỏ ж Ủ ụ ỹ ố ö Ẽ Γ ụ ố Ủ ẩ



6 8 ẩ ö ж ề ỵ ố ễ Ẽ ī ố ö ố ж ề ч Θ ố ể ụ σ ẩ Ỹ ī ж Ỹ ж ề Y ề ж ö γ ề Y Ỹ ể ố Ủ ẩ

- ‘ **ö ж ề ỵ ố ễ Ẽ ī ố ö ố ж ề ч Θ ố ể ụ σ** ỏ Ẽ ủ ỵ μ Θ Ẽ ố ỹ Ỹ σ ể ỹ ễ ẩ Θ ỹ ж ī ö ụ Ỏ ж Ỹ ỹ Ẽ ỹ ж Θ ể ж ể Y Ỹ ж ủ ж Ủ Ỹ ỵ ж Ó ể ж Ỹ ỹ ụ ễ ж Ỏ ố ỵ ụ ễ Y ễ ж Ó ố ễ ж Θ ể γ Ỹ ụ ỵ γ ể ж ч ễ ố ể γ Π ụ ỹ γ ж Π ụ ề ủ ч ể Ẽ ж ề ч Θ Ẽ μ Σ ố Ủ ể Y ề ж ö ủ ố ẩ ể ж ể ễ ố Ỹ ỹ ố Ỹ ζ ỹ ụ ễ ī ж ỹ ố ö σ ỹ γ ж ề ж ề Σ Ẽ μ Σ ч μ Ỹ ī ж Ỹ ж ề ể ж Ỹ ỹ γ ễ ж Ỏ ố ỵ ụ ẩ Ш ч ề ж ễ ж ể ж ö σ ẩ Ẽ ỵ ễ ж ö ụ ỹ ễ Ỹ ỵ ụ Π ủ ж ễ Ỹ ụ ỵ γ ể ж ī ж Ỏ ễ ж ể σ ố ỹ Ỹ σ ī ж Ỏ ж ề ч Θ Ẽ μ Σ ч μ ể Y ề ж ö ủ ч ẩ
  - ˜ ж ö ж Π ụ ễ ī ö ụ ễ ố ö ж ễ σ ể ỵ σ ố ỹ Ỹ σ ỏ Ẽ Ỏ Ẽ Θ Ẽ ể ж Ỹ Ỹ ỹ Ẽ ể ж ể ỵ ố ể ụ σ ễ ể ж ễ ж Θ ỵ ố Ẽ ī ж Ỏ Ẽ ể ể Y ễ ỹ ж Θ ủ Ẽ ễ ẩ x ỏ ể ố Ỹ ỹ ể ж ẩ Θ ỹ ж ố Ỹ ỵ ụ Ỏ Ẽ ể Y ỏ ể Ẽ Θ ố ể ụ σ Ř ч ể ủ Γ ụ ụ ể k ỹ ж Θ ủ Ẽ ẩ ỹ ỹ ж ī ж Ỹ ỹ ö ж ụ ỹ γ ễ ể ж ễ ж Θ ỵ ố ể Ỹ ỹ ố ī ố ể ụ k, ủ ж ỹ ж ö Y Ủ ī ö ж Ủ Ỏ Ỳ ỹ ỹ ж Θ ể ж Θ ố ö ố ở Ỏ Ẽ ể ể Y ố ỹ ж Θ ủ ẩ Ш ỹ Ẽ ủ ж ễ Ỹ ỵ ч Θ Ẽ ố ỵ ụ ể ố Ủ ể Ẽ σ ễ ж Ỏ ố ỵ γ ể Ẽ Ỏ ī ö ụ ỏ ể Ẽ ủ Ẽ ễ ụ $(1, x, x^2, x^3, \ldots x^k)$ Ỹ ễ ж Ó ố ỹ Ỏ ж Ỹ ỵ Θ γ ể ч ỵ ố ể ж Ủ ж Π ụ ề ủ ẩ Н ь ж ẩ ж Θ ố ể ụ Ỏ ể ж ẩ Θ ỹ ж ố Ỹ ỵ ụ ể ủ Ẽ Θ ố Ỹ ỹ ể ố ẩ ụ Ỹ ỹ ж Θ ể ụ ủ Ẽ ỹ ж Θ ủ ẩ ể ở σ ỹ γ ễ ể ж ễ ж Θ ỵ ố ể ể ụ ể Ẽ ố ỹ ж ề Σ ч μ ỏ Ẽ ể ụ Ỹ ụ ễ ж Ỹ ỹ γ ẩ Ẽ ī ж Ỏ Ỹ ỹ ö Ẽ ụ ể Ẽ ố ỹ Ỹ σ ī ж Ỏ Π ч ễ ẩ Ẽ ủ Ẽ σ ễ ж Ỏ ố ỵ γ ề ч Ỏ ố ỹ ụ ễ ố ỹ γ ж Θ ể γ ề ж ỵ γ Π ч μ ж Π ụ ề ủ ч ể Ẽ ỹ ж Θ ủ Ẽ ể Ỹ ễ ố ể ố ö ụ ö ж ể Ẽ ể Ẽ ể Y ř ỹ Ẽ ủ ụ ễ Ó ố Ỹ ī ж Ỹ ж ề ж ễ ẩ ể ể ж ể ố ể ř ж Ỏ σ Σ ụ ř ể ж ề ч Θ Ẽ μ Σ ч μ ể Y ề ж ö ủ ч ẩ
- ˝ ỹ ж ề Y ụ ỏ ề ố Ó Ẽ ỹ γ ī ố ö ố ж ề ч Θ ố ể ụ σ ể ễ Ẽ Π ụ ể ể ж ễ ж ề ч Θ ố ể ụ ụ ī ö Ẽ ủ ỹ ụ ủ ч ố ỹ Ỹ σ ể ố Ỹ ủ ж ỵ γ ủ ж ö Ẽ ỏ ể Y ř ễ ố ỹ ж Ỏ ж ể ẩ H
  - *ю ã ã ũ õ Æ ẽ ũ õ ð ũ ã Ě ủ ỵ Q ũ ỵ Ě ж Ò ủ ỵ ế ể ủ ố* ˚ H Ẽ ī ö ụ ễ ố ö ẩ ễ Y ễ ж Ó ố ễ ж ễ ö Ẽ ể ụ Θ ụ ể Ẽ ỹ γ ễ ỹ ч ề ụ ể ч Ỏ ố ö ố ể γ ố ể Θ ỹ ж ề Y ч ễ ố ể γ Π ụ ỹ γ ể ж ỏ ễ ж Ó ể ж Ỹ ỹ γ ī ố ö ố ж ề ч Θ ố ể ụ σ ẩ a ö ч ễ ж Ủ ī ö ụ ễ ố ö ẩ A ö ố ễ ч ỵ σ ö ụ ỏ Ẽ Γ ụ σ ỵ ụ ể ố Ủ ể Y ř ễ ж Ỏ ố ỵ ố Ủ ẩ Ш Ř ч ể ủ Γ ụ μ Ỏ ж ề Ẽ ể ỵ σ ố ỹ Ỹ σ Π ỹ ö Ẽ Ř ỏ Ẽ Ỹ ỵ ụ Π ủ ж ễ ề ж ỵ γ Π ụ ố ể ố Ỹ Ẽ ễ ж Ỏ ố ỵ ụ ẩ ủ ж ỹ ж ö Y ố σ ể ỵ σ μ ỹ Ỹ σ ī ö ụ ỏ ể Ẽ ủ ж ễ ī ố ö ố ж ề ч Θ ố ể ụ σ ẩ
  - *M ể ủ ể ũ ố ể ả Z ả ỵ ă ỏ ố ũ ã ũ ể Ě ỏ ỵ ủ Q ỏ ỵ Æ ã Ě ж ể Ô* ж ề ч Θ ố ể ụ ố ẩ ể Ẽ ỵ ụ Ỏ Ẽ Γ ụ σ ụ ỹ ố Ỹ ỹ H Ỹ ễ ẩ ī ö ố Ò ī ч ể ủ ỹ ı ẩ
  - *ĭ ă ỵ Ě Ě ˝ ả ã ủ ố ế ã ỹ ố ð* ˛ " ố Ò ī ж ễ ж Σ γ μ ī ж ỵ ч Θ Ẽ ố ỹ Ỹ σ ề ж ỵ ố ố ч Ỹ ỹ ж Ủ Θ ụ ể Ẽ σ ж Γ ố ể ủ Ẽ ủ Ẽ Θ ố Ỹ ỹ ể Ẽ Ẽ ỵ ễ ж ö ụ ỹ ễ Ẽ ẩ ố Ỹ ỵ ụ ễ ố Ỏ ж ī ч Ỹ ủ Ẽ ỹ γ ч ỹ ố Θ ố ủ Ỏ Ẽ ể ễ Y ř ẩ ж Ỏ ö ж ề ễ ố ố Ỹ ễ ī ö ố Ỏ ẩ ī ч ể ủ ỹ ẩ
  - H ь ố ủ ж ỹ ж ö Y ố Ỏ ö ч ễ ụ ố ī ö ụ Ỏ ễ Y ' H
    - T ể Ẽ ỵ ụ ở ī ö ụ ỏ ể Ẽ ủ ж ể ể ể ỹ ẩ ẩ ẩ

6 + ňĵ ŏẽÿ ŋ ŏöŋ ŏẽÿ ŋ ŏöŋ ŏẽÿ ŋ ŏöŋ

ħ ÒĕĒźżīöţ ţŸīżŷŷòżĕĕĒĕţţ ^ ŃţÜÕ" ĻĦĥ īżŸŷŷŏ Ÿŷżœ T ē Y Řţźŷţöčŏē Ÿöŏ Ò ĕŏŏ ţ
Ò ţŸīŏöŸţţ ŏĕżĕŷŸřżÒĒĥŒŷżīżŒŷţ īżŷĕżŸŷŷţţ öŏ Ĕŏŷ ζŷč īöżĕŷŏē čĥ

- k Ĵ ĤĴ" ŏńA Ÿŷżźŏ Ò öżĕĕżźŏ Ÿŏŷţĥźżŷżöŷ Ò öŏĕčŷœöţòčŏŷ ĕĒĒčē żÒŏŷŷĥ čŸŷŷē Ĕŷöţ ĒĒ X A ĕřżÒĕŷŏ ÒĒĕĕŷŏ ζŷżĕżŸŷżœĥŠżĕöŏĕē œÒŷœźĒÓÒżĕżζŷŏĕē ŏĕŷĒē Ĕŷöţ ĒŷÒĕčŷţē ŏĕżŸ ÒĕöĒĕĕŏŏ ĕŷŷżźŏżŷżöŷĕżźŏżŷŷţţ p." żżŷĕŏŷŸŷĕŏĕĕżĥŸ ĕŏöżœŷĕżźŏżŷŷţţ $1-p$ ē Y ĕţŒŏĕżĕŏ Ò ŏŷĒŏē Ÿζŷŏĕē ŏĕŷżĕē ē Ĕŷöţ Ēŷŷŷĕĕŷœŏĕē ŏĕżĕ ĕřżÒŏ Ÿŷżœĥ
  - " čŷŷ Ÿŷżœĕ ŷżĕĥŒŷżē Y żĕčŒĒŏē ĕżŷŏŏ Ÿŷ Ĕĕčţ ē żÒŏŷŷ ŷźĒźŷźŸżöĕĕĒŷŏŷĕĕżţĕĕżöţöčŏē ĕŏźŷŷżöčţ ŷŏŏčţ ŷŏŏčŷŷ ŷŏŏĕĕĒŷŏŷĕĕżŷŏĕē ĕżĕŷŷŏŷŏŏŷŷŏŷŷŏŷŷ

    - ħ ÒĕĒ ŸÿöżżůĒ ē ĒÿöцΓΥ ı żÒцĕ ζцőē őĕÿ żĕöčΘĒμ Σ őŮ ěΥĕżöůцŋ̂

- – **őцőěĒσ īőöőőē őĕěĕĒσ** ’A Θцŷцżцццē őÿůĒ ůцĒŸŸĒŋŸÿĒěσΣ ĒσŸσ ě Ÿżżÿěőÿŷěůцő żÒĕżĕē ч żĕTőůÿčŋ̂

- ‘ **öцòĕĒůr** ’A ř ĒöĒůÿőöцŸÿцůĒ żĕTőůÿĒ нцμ ĕżŮůčŸżΘőůцĕŘżöē ĒΓццŋ̂ůżÿżöΥŮĕĒē ÒżŸÿčīőĕ żĕ żĕTőůÿőıŋ̂ öцòĕĒůцē żĕčÿ ĕΥÿγ ΘцŷżěΥē ц (ℝ)ŋ̂ůĒÿőēżöцĒγĕΥē ц нцò Řцůŷцöżěĕĕżēż òĒöĒĕőő ē ĕżÓőŸÿěĒı čīżöσÒżΘőĕĕΥē цŋ̂цццē ĕцĕĒöĕΥē цŋ̂

    - ‘ żŸčÿцī īöцòĕĒůżē σěγσőÿŸσ цμĕżŮůčŸżΘőůцĕŘżöē ĒΓццī żĕ żĕTőůÿőŋ̂

    - ‘ öőÒīżÿĒēĒőÿŸσŋ̂ΘÿżěŸő żĕTőůÿΥ ě öĒē ůĒŕ żÒĕżŮòĒÒĒΘцī żīцŸΥěĒμÿŸσżÒĕцē ц ц ÿőē ц Ó́ő īöцòĕĒůĒē цŋ̂

- **ЛżÒőγγ** нĒÿēżöцÿē ı ’A żÿżĕöĒÓ́őĕцő цò īöżŸÿöĒĕŸÿěĒ īöцòĕĒůżě ě ē ĕżÓőŸÿěżΓőγőěżŮ īőöőē őĕĕżŮнżĕTőůÿ ōĭ Γőγőěĕσ īőöőē őĕĕĒσıŋ̂

- ‘ **čĕůΓцżĕĒÿ żΠцĕůцī** ’A ŘčĕůΓцżĕĕÿ ĕĒ ē ĕżÓőŸÿěő Ēÿēżöцÿē żěŋ̂òĒěцŸσΣ цŮżÿ żĕčΘĒμ Σ őŮ ěΥĕżöůцŋ̂

    - ΓěγσőÿŸσ ř ĒöĒůÿőöцŸÿцůżŮůĒΘőŸÿěĒ Ēÿēżöцÿē ĒΗΘőē ĕżÿγΠő öĒòĕцΓĒ ē őÓ́Òč īöőÒŸůĒòĒĕцσē ц Ēÿēżöцÿē Ē Òγσżĕ TőůÿżěĕżĕčΘĒμ Σ őŮ ěΥĕżöůцī ц цř ÿĒöēőÿĒē цŋ̂ÿőē ĕżÿγΠő òĕĒΘőĕцő ŘčĕůΓцżĕĕÿ Ē żΠцĕůцŋ̂

- **ħĕčΘőĕцő** ’A īöżΓőŸŸ īżÒĕżöĒ īöĒē őÿöżěŋ̂żÿ ůżÿżöΥř òĒěцŸцÿ Ēÿēżöцÿē ŋ̂ÿĒůцē żĕöĒòżē ŋ̂ ΘÿżĕΥ ē цĕцē цòцöżěĒÿγ ŘčĕůΓцżĕĕÿ żΠцĕůцŋ̂

    - ¸ ŋ̂řŋ̂īőöőÒżĕčΘőĕцőē ŘцůŷцöčőÿŸσ ĕőůżÿżöżő *Ĕěůřěężĕγ ãůãÿă õżůγě* ěĕčÿöцůżÿżöżĕē ē żÒőγцżÿцΘĒμÿŸσ Òöčē żÿ Òöčē ĒÿцΠγòĕĒΘőĕцē ц īĒöĒē őÿöżěŋ̂

7 ŋ̂ ÿżÿĒůżő ŸěżŮŸÿěżāāŋ̂ŋ̂ā

Æ̧l XḤXę̧l ę̧ṁ ę̧l ÆEXę̧mā̧ ĒĒi k ā̧kmḥĒ̄ő̄mXl ā̧ĕőòĒěцŸцē Yő ц żÒцĕĒůżěżöĒŸīöőÒőγőĕĕYőŋ̂
ШĕĒΠцĕĕżē żĕčΘőĕцц żĕYΘĕżīöőÒīżÿĒēĒőÿŸσŋ̂ΘÿżżĕčΘĒμ Σ ĒσěYĕżöůĒěŸőĒěÒĒčÒżěγőÿěżöσőÿ ζÿżē ч ŸěżŮŸÿěčŋ̂Λÿżòĕ ĒΘцÿ’H

- ШŸő żĕTőůÿΥ ěYĕżöůцī ĕőòĒěцŸцē Y Òöčē Ÿ Òöčē żē нÿŋ̂řŋ̂цř ē żÓĕżŸΘцÿĒÿγ ĕőòĒěцŸцē Yē ц ŸγčΘĒŮĕYē ц ěőůÿżöĒē цı ’Ā̂ ζďζĵ ‘ĵĵ‘
- ШŸő żĕTőůÿΥ ěYĕżöůцī żÒцĕĒůżěżöĒŸīöőÒőγőĕΥ нÿŋ̂řŋ̂ζÿцī ŸγčΘĒŮĕYő ěőůÿżöĒ цē őμÿ żÒцĕĒůżěżő öĒŸīöőÒőγőĕцőı ”ĂζĵĵĂ‘ ĜĜĂĂĂ

[+] ญ ยังยืนยันข้อมูลการเข้าใช้งานระบบ

- **ปัญหาที่ต้องแก้ไข:** ในขั้นตอนที่ผู้ใช้งานกรอกชื่อผู้ใช้งานและรหัสผ่านเข้าใช้งานระบบ ระบบจะทำการตรวจสอบข้อมูลที่ได้รับกับฐานข้อมูลผู้ใช้งานที่ลงทะเบียนไว้ในระบบ
- **ฟังก์ชันที่ใช้ในการตรวจสอบข้อมูลผู้ใช้งานเข้าใช้งานระบบ:** ในขั้นตอนที่ผู้ใช้งานกรอกชื่อผู้ใช้งานและรหัสผ่านเข้าใช้งานระบบ ระบบจะทำการตรวจสอบข้อมูลที่ได้รับกับฐานข้อมูลผู้ใช้งานที่ลงทะเบียนไว้ในระบบ

) ๆ ๆ ๆ

öőẽÿụơöụỏẽĞụơỎőỵẽőÿ *ỵẋằỵă ìăõõữẫỏỵẳ* ỵỉủỉẋẽ ỉểỴềụöẽőỵỉỉỸẽẽỴőő ểẽÓẽỴőő ïöụỏẽẽửụ ụ ẋỸỹẽểỵơőỵ ụở ể ẽẽ ẋỎőỵụỉỉ

- t Ề ḳṇ̃ ÔXṇ̃H $L = \frac{1}{n}||Xw - Y||_2^2 + \lambda_1^2||w||_1 + \lambda_2^2||w||_2^2$ ’A ửẋẽề ềụẽẽĞụơ ‡⁵ ụ ‡⁶ öőẽÿụơöụỏẽĞụụỉỉ
  - " ỵẽẽẽőẽẽẋőỉỉỎẋềẽểỵơőẽẽẋő ửụỸỉẋỎẽẋẽẽ ÿ ỈỉÿẽửĞụẋẽẽỵÿ ẋỉ Ụềửụ ẽẽỎỴểẽőỹỸơ *ăểẫỉủõăõõẫỳỵăỵửỵ* ểẽẋẽẽẋỎẽẋẽẋïöụềẽểỵơỹỹ ẽőỉ ỹẋỵỹửẋửỸöőỎẽőửểẽỎöẽỹụỦẽẋẽẋỎẋỉ ụềửőỉỉửẽửẽẽẽẽẽẽẽẽẽỵẽẽỵỉỉỹẽẋẽẋẽẋẽỵỉỉẽẽẽẋẽẋẽẋẽẋẽỵỉỉẽẽẽẋẽẋẽẋẽẋẽỵỉỉẽẽẽẋ

- ôÔ ňô ĦĶXęXnj mĦXı ’A ůżŷụŒŏŸŷểżżềTốůŷżểīżŏụŷụểểżểżůŷẾŸŸẾňůżŷżöYố ůŷẾŸŸụŘụůẾÿżö ểẾŏểẾŷ ểŏểẾÿụểểYểụňĵ

		True condition	
	Total population	Condition positive	Condition negative
Predicted condition	Predicted condition positive	**True positive**	**False positive**, Type I error
	Predicted condition negative	**False negative**, Type II error	**True negative**

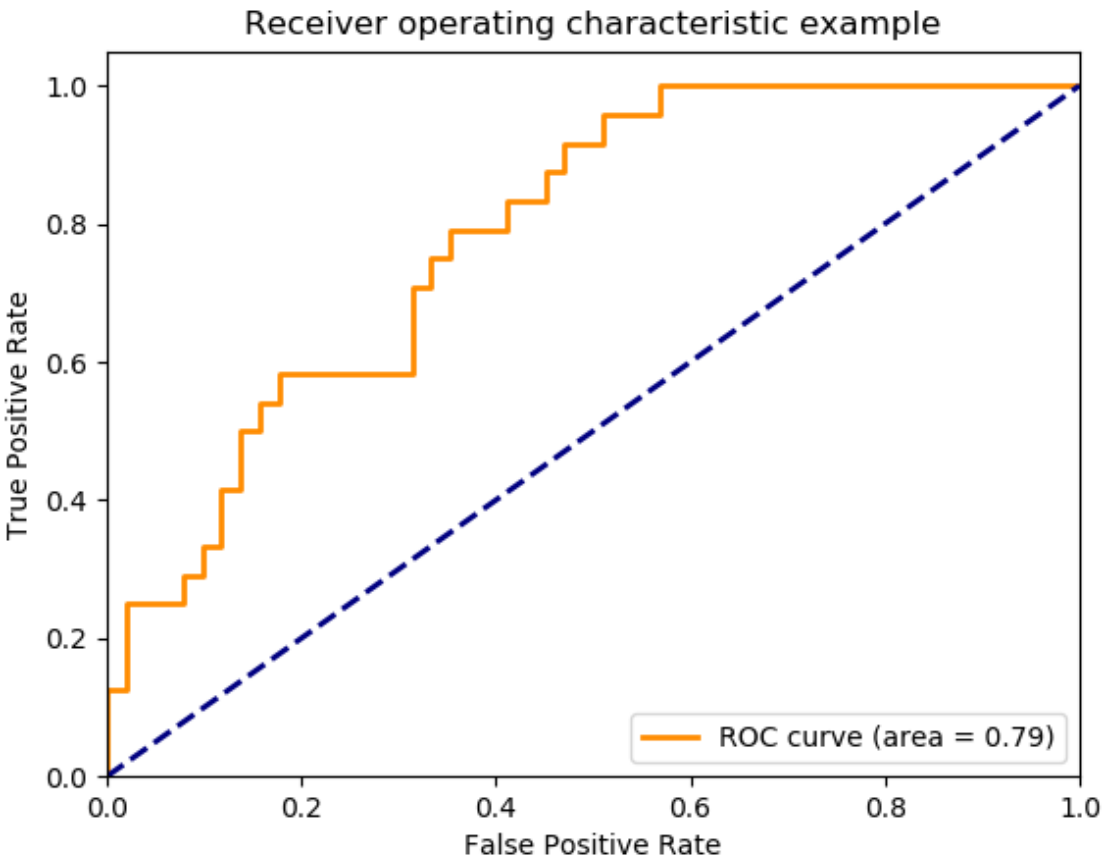
[5] [6] ňĵ ÿżŷẾůżố Ė˙ b ŏ2ń b đ ə Ẽử īżŸÿöżụÿγ Ė˙ b ŏůröụểŷμ đ

- **Ė˙ bŏ ŏĦ X** ’A ẽöẾŘụử ểżŸơř ôĒĖ ňô ĦĶXẾ” ķāĦXĖ mXňżŸγ xı ụ ĶẾĖ ňĶĵŏXẾ” ķāĦXĖ mXňżŸγ yı ňĵ

$$\mathrm{FPR} = \frac{FP}{FP + TN},\ \mathrm{TPR} = \frac{TP}{TP + FN} (= \mathrm{Recall})$$

  - ôẾĖ ’A Ỏżŷσểố īöẾểụyγểżŷểẾỎẾểểYř żềTốůŷżểểố ểẾŷụểểżểżůŷẾŸŸẾňĵ
  - ĶẾĖ ’A Ỏżŷσ īöẾểụyγểżŷểẾỎẾểểYř żềTốůŷżểểīżŏụŷụểểżểżůŷẾŸŸẾňĵ
  - x ŸīżŷγỏểöŷŸσỎŷσůŷẾŸŸụŘụůẾÿżöżểňīöốỎŸůẾỏYểẾμΣụř ểốöżoŷểżŸŷγ īżŏụŷụểểżểżůŷẾŸŸẾ ňżểYŒểżůŷẾŸŸ Ÿểốÿůżů $+1$ı ňĵ
  - ‘ **żŸŷöżốểụố’H**
    - ‘ żŸŷố īżŷểŒốểụσểốöżoŷểżŸŷốỦỎŷσżểTốůŷżểểYểụöẾốÿŸσẽöẾểụΓẾňůĦĶ’ ĦĖı ỎŷσểốöżoŷểżŸŷụňĵвŸŷụểốöżoŷểżŸŷγżểTốůŷẾểYΠố ζŷżỦẽöẾểụΓYňżểżŷểżŸụÿŸσůīżŏụŷụểểżểểể

- ‘ öцẽ ốö Ė̇ b ở̈ử̈öцể̈ӝỬ̈H



[5 7] ŋ̂p Ẽīцп цÿố Ř ӵể̈ử̈Г цӝể̈Ẽў у ӝẽ̈цŸ̈ÿцѲố Ÿ̈ử̈ӝỬ̈ öố ẽ̈öố Ÿ̈Ÿ̈ццŋ̂э Ẽử̈ ӝể̈ Ÿ̈ể̈ơở̈Ẽể̈ Ÿ̈ ẽ̈ ốÿӝỎӝẽ̈ ẽ̈ Ẽử̈Ÿ̈цẽ̈ ӵẽ̈ Ẽ̈ īöẽ̈ể̈ỎӝīӝỎӝề̈цơđ

- ‿ ӵể̈ử̈Г цӝể̈Ẽў ӝПцề̈ửц Ỏцơ у ӝẽ̈ŋ̂öố ẽ̈öố Ÿ̈Ÿ̈цц нŧ” nŧ” ḳḳı'H

$$L = \sum_{i=1}^{n} \log(1 + \exp(-y_i \cdot w^T x_i))$$

- Лцể̈цẽ̈ цở̈Ẽ̈Г цơ ζÿӝẽ̈ӝ Ř ӵể̈ử̈Г цӝể̈ẼўẼ ζửể̈цể̈Ẽўố ể̈ÿể̈Ẽ̈ ẽ̈ Ẽử̈Ÿ̈цẽ̈ цở̈Ẽ̈Г ццц īöẽ̈ể̈ỎӝīӝỎӝề̈цơ ể̈Ÿề̈ӝöửц'H

$$\text{Likelihood} = \prod_{i=1}^{n} P(x_i \text{ принадлежит классу } y_i)$$

- a ӝửẽ̈ở̈Ẽ̈ÿố у γŸ̈ÿể̈ӝζÿӝẽ̈ӝ ӵÿể̈ố öỎỏố ể̆цơ ẽ̈ ӝỎể̈ӝīӝŸ̈ẽ̈ ӝÿöố ÿγ ÿ̆ӵÿ̆ŋ̂

[5 8] ŋ̂ħ īцп цÿố Ỏẽẽỏể̆ Y Ử̈ Ẽўẽ̈ӝöцÿẽ̈ ӝề̈ ӵѲố ể̆цơ öố П Ẽμ Σ ốẽ̈ӝ Ỏố öố ể̈ể̈Ẽ̈ŋ̂

- ‘ ӝŸ̈ÿöӝố ể̆цố ể̆Ẽ̈Ѳцể̆Ẽ̈ố ÿ̈Ÿơ Ÿ̈ ửӝöể̆ơŋ̂Ш̈ửẽ̈Ỏ̈ỎӝỬ̈ể̆ӝỎố Ỏố öố ể̈ể̈Ẽ̈ ẽ̈ Y Ỏố у цẽ̈ ể̈Y ề̈ӝöửӵ ể̆Ẽ̈ Ỏể̈ố Ѳ̈Ẽ̈Ÿ̈цī ӝ Ÿ̈ цố Ỏӵμ Σ ốẽ̈ ӵ īöцể̆Г цīӵ'Hбŷ̈ццī ơể̆Ẽ̈ố ể̆цố j ở̈ӝẽ̈ӝ īöцơể̆Ẽ̈ửẽ̈ ӝề̈ Tốửÿ̈Ẽ̈ x ẽ̈ ốể̆γП ố Ѳố ẽ̈ ẽ̈öẽ̈ể̆цГ Ẽ̈ t, ÿӝ ӝề̈ Tốửÿ̈ īӝīẽ̈Ỏ̈Ẽ̈ố ÿ̈ ể̈ īöẽ̈ể̆ӝố īӝỎ̈Ỏố öố ể̈ể̈ӝŋ̂цể̆Ẽ̈Ѳố ’A ể̈ ує̆ể̆ӝố ŋ̂ цŸ̈ує̆ j цī t ể̆ố ӝề̈ř ӝỎцẽ̈ ӝīӝỎӝề̈öẽ̈ÿγ Ỏцơ ửẽ̈Ỏ̈ỎӝỬ̈ể̆ӝỎ̈Yŋ̂ Ẽ̈ửŋ̂ể̈Yề̈ӝöửẽ̈ ể̈ ÿố ửӵΣ ốЎ̈ể̆ӝỎố öẽ̈ở̈Ỏố у цÿŸ̈ơ ể̆Ẽ̈ Ỏể̈ố īӝỎể̈Yề̈ӝöửцī ’A ӝỎể̆Ẽ̈ ӝửẽ̈Ỏ̈ố ÿ̈Ơ ể̈ īöẽ̈ể̆ӝẽ̈ īӝỎ̈Ỏố öố ể̈ố ŋ̂Ỏöӵẽ̈Ẽơ ể̈ ує̆ể̆ӝẽ̈ ŋ̂

- pẽ̈ẽ̈ ố ÿцẽ̈ ŋ̂Ѳÿӝ īöцơể̆Ẽ̈ửӝể̈ цī ӝề̈ Tốửÿ̈ӝể̈ ӵẼ̈ ơể̆Ẽ̈Ѳцÿ̈ цī ơể̆Ẽ̈Ѳố ể̆цỦ̈ửẽ̈Ỏ̈Ỏӝẽ̈ӝ īöцơể̆Ẽ̈ửẽ̈ī ’A ửӝể̆ố Ѳể̆ӝố Ѳцŷ̈ує̆ŋ̂ ӝẽ̈Ỏẽ̈ ẽ̈ Y ẽ̈ ӝỎố ẽ̈ īöố öố ề̈öẽ̈ÿγ ể̈Ÿ̈ố ể̈ӝơể̆ ӝỎể̆Yố īẽ̈öY (j, t), ửӝÿӝöYố Ỏố у ơÿ̈ ể̈Yề̈ӝöửӵ ể̆Ẽ̈ Ỏể̈ố Ѳ̈Ẽ̈Ÿ̈цŋ̂ħ īöố Ỏố у цẽ̈ īöцể̆Г цīŋ̂īӝửӝÿӝöӝẽ̈ ӵ ề̈ӵỎố ẽ̈ Ÿ̈öẽ̈ể̆цể̆Ẽ̈ÿγ ζÿцī īẽ̈öY ẽ̈ ốỎ̈Ỏӵ Ÿ̈ӝề̈ӝỬ̈ŋ̂

- Шеőőõõõ

öőΠĒμΣőő Ȍöőőểжčèчȍőÿ īöőȎŸůĒȍYểĒÿγ ē őÿůч ζÿжĒжжĒTőůÿĒŕÿĒůĒůжĒTőůÿжểŸȎöчĒůē ů ē őÿůĒĒů ů ểůŸÿőŏőÿčŕ жőŸÿγȎöőőểжīöжŸÿжȍĒīжĒŏůÿŏÿčĒΠч жčèчȎĒμΣчμểYčжöůч ů īжζÿжĒч жΠůčůĒčèчȍőÿŏčůőểĒőŕ

5 - ŕ ÿжÿĒůжő 6 njnjäęnjā

- **^ njnjäęnj** н^"" nkmŕHŖ njnjĮXnj mäęnjı ΆĀ ē őÿжȎĒčŸĒē čÿůöжểĒčůσĒÿēжöůÿē жểŕůжÿжöYŮ ȍĒůÿμȎĒőÿŸσểжčèчȎőčůů N čĒȍжểYř ē жȎőÿőŮčĒčèчÿŸÿöĒīůöжểĒčYř ểYčжöůĒř ů чŸöőȎčőčůů ůř жÿểőÿĒŕ
  - вŸÿů īöőȎīжÿжȎůÿγŕȎÿжжΠůčůů čĒȍжểYř Ēÿēжöůÿē жểčőŸůжöőőÿůöжểĒčY нȎÿжŕ жȎőểůȎčжŕčőīöĒểȎĒŕÿĒůĒůжčèчȎĒμΣůő ểYčжöůů čèчŸÿöĒīůöжểĒÿůŸγůȍжčΣőēж ē čжȎőŸÿểĒжčTőůÿжčŕŕů ůř ȎůŸīőöŸůůů öĒččY ē őȎȎч Ÿжčжŕŕÿжčζččůčē čē čőčγΠĒőÿ öĒȍčöжŸčĒȍжểжēжĒÿēжöůÿē Ē ểΝ öĒȍŕ

  - ŕŕŕčζččůčē čőčē čőčσőÿ 6 ā ķ čĒȍжểYř Ēÿēжöůÿē жểŕčжčē čőčγΠĒőÿ ůř ȎůŸīőöŸůμŕ
    ‘ жζÿжē ч жč Ÿůÿγčжчčőčůů ȎůểĒőÿ ůĒȎőŸÿểжȎőöőểγőểŕůжÿжöYő жčÿĒȎĒμÿ čůȍůůē Ÿē őΣőčůőē ŕčжểYŸжůůē öĒȍčöжŸжē ŕ

5 (ŕ ÿжÿĒůжő ŸÿůчȎĒŮčYŮůőŸā ő ē жč жÿůÿůȎĒőÿŸσжÿ čζččůčē čĒ čĒȍ öőΠĒμΣůē ů Ȍőöőểγσē ůā

Ė ęl "ň ô"ĮXkmн ůчȎĒŮčYŮůőŸı ΆĀ ē жȎőÿγŕσểÿσμΣĒσŸσĒčŸĒē čÿőē ȌőöőểγőểŕůжÿжöYŮ жčèчȎĒőÿŸσŸīжē жΣγμ čζččůčē čĒů ē őÿжȎĒ ŸÿůчȎĒŮčYř īжȎīöжŸÿöĒčŸÿểжȎčжểöőē őččжŕŕ ŕŕŕ ůĒȍŕȎYŮčĒȍжểYŮĒÿēжöůÿē жčèчȎĒőÿŸσčĒčèчÿŸÿöĒīůöжểĒččжŮīжȎểYčжöůő ŸжŸÿůчȎĒŮčж ểYčöĒččYē īжȎē čжȍőŸÿểжē īöůȍčĒůжểŕ

- **ħ ÿÿůчȎůőжÿčζččůčē čĒ čőőΠĒμΣůř Ȍőöőểγσř** ȍĒůÿμȎĒőÿŸσểůŸīжÿγȍжểĒčůů ē őÿжȎĒ ŸÿůчȎĒŮčYř īжȎīöжŸÿöĒčŸÿểȎÿσжčèчȎőčůσůĒȍŕȎжēжčèȍжểжēжĒÿēжöůÿē Ēŕ
  - *đ'Ĩ ùÊ ĎĮ Ğů' Ţ/ ŢĴdů* нĖgÎ ı ΆĀ ē őÿжȎŸÿůчȎĒŮčYř īжȎīöжŸÿöĒčŸÿểŕTčĒÿжčůȎčжčζččůčē ččчжčèчȎĒőē ůĒȍŕȎYŮčĒȍжểYŮĒÿēжöůÿē ůŸīжÿγȍчσÿжÿγůжŸÿůчȎĒŮčжő īжȎē čжȍőŸÿểжīöůȍčĒůжểŕ öů ÿĒůжē īжȎřжȎő Ēÿēжöůÿē Y ŸÿĒčжểσÿŸσőΣ Ȍčèжÿőő čőŸůжöőÿůöжểĒččYē ůŕ
- ŕůŕŕȎőöőểγσжȍőčγ ŸÿĒčèжŸůжöőÿůöжểĒčY Ȍöчē Ÿ Ȍöчēжč īöů ÿĒůжē īжȎřжȎőŕöőĒÿγčYŮ öĒȍčöжŸÿůÿγčőő īöůčèÿůȍĒőÿŸσůжΓőčůő $\frac{\text{Разброс базового алгоритма}}{N}$.
- ĒůжŮē őÿжȎжčèчȍőčůσжȍőčγζŘŘőůÿůčőčē ůē čőččжȎÿσȎőöőểγőểŕīжÿжē ч ȍÿжȎőöőểγσÿőĒůжīőőőжčèчȍĒμÿŸσůŕůĒůŸÿőȎŸÿểůőŕÿčèÿĒȎĒμÿ čůȍůůē Ÿē őΣőčůőē ů čểYŸжůůē öĒȍčöжŸжē ŕ
  ē őčγΠĒσöĒȍčöжŸŸīжē жΣγμ čζččůčē čĒů ĖgÎ ŕē Y īжÿчȍĒőē ē жȎőÿγŸčůȍůůē ů Ÿē őΣőčůőē ů öĒȍčöжŸжē ŕ

5) ŕэĒůểē čöĒȎůőčÿčжē čчŸÿůččőő жčèчȍĒμÿŸσčĒȍжểYő Ēÿēжöůÿē Yā

ЧчŸÿůčē ΆĀ ē őÿжȎĒčŸĒē čÿůöжểĒčůσĒÿēжöůÿē ÿжểŕ
б ểểδ ầпĒжőũãậ

9 ŕЧчȎőē Ÿÿöжůÿγ ůÿжčжểYŮĒÿēжöůÿē ůĒůŸчē ē ч čĒȍжểYř Ēÿēжöůÿē жểůȍжȎčжčжŸőē őŮŸÿểĒŕ

$$a_N(x) = \sum_{i=1}^{N} b_i(x)$$

- ęůŸŲöčőē ĕőůżÿżöYŮĕĒŒĒŲγĕYŮĒŲężöůÿē $b_0(x)$.
- ężĒŐŎYŮŸŲőŎčμΣůŮĒŲężöůÿē ŸÿöżůÿŸσÿĒůęŒÿżĕYůŲĒöĒęůÿγżĕΠůĕůčŋčŐőē īżŸÿöżőĕĕżŮůżĕē īżöůΓůμ ŋčē őĕγΠĒσĕőůżÿżöčμ ŘčĕůΓůμ īżÿőöγ‡ı'H

$$b_N(x) = \arg\min_b L(x, a_{N-1}(x) + b(x))$$
$$a_N(x) = a_{N-1}(x) + b_N(x)$$

ŠŋĒŎůőĕÿĕYŮĕčŸÿůĕē'A ĒężöůÿēŎęσīżŸÿöżőĕůσĕčŸÿůĕęĒęęĒęůŸīżÿγöčμΣůŮęöĒŎůőĕÿYŎęσčē őĕγΠőĕůσżΠůĕůůęęĒ
*ŋůęăőżŭ ăăăęőęęŭżÿęăÿ ăīĒżőŭăăęĒ*H

- ęŠŸĕůčŸÿĒĕÿĕYŮĕĒöżĕęŸŮĕĒżöůÿē ŋŸöőŎĕőő ęöĒŎĒŒő öőęöőŸŸůęęęŎżē ůĕůöčμΣůŮ ůęĒŸŸęęYĕżöżŒĕYő ęőöżσÿĕżŸÿů ęöĒŎĒŒő ůęĒŸŸůŘůůĒΓůęīęęę
- ęŠŸŒůŸęőē ęőůÿżö ĒĕÿůęöĒŎůőĕÿĒ ŘčĕůΓůů īżÿőöγ īöżůöężŎĕčμ īżężöęöĒΣĒőē Yē öĕĒŒőĕůőē čŐőē żĕčŒőĕĕżŮŒĒŸÿů ĒĕŸĒē ĕęσĕĒ żĕčŒĒμΣőŮęYĕżöůőęę
- ęÿöżůē Ēężöůÿē öőΠĒμΣůŮöĒŎĒŒčī öőęöőŸŸůęęęöĒŒőŸÿęőŐ ŘčĕůΓůů żΠůĕůů ůŸīżÿγöčσ Î gt 'H$b_N(x) = \arg\min_b \frac{1}{l}\sum_{i=1}^{l}(b(x_i) - s_i)^2$
  - żőŸÿγēYůΣőē ÿĒůżŮĒężöůÿē b_N, ŒÿżĕYīöůĕĒęęęőĕżůīżŸÿöżőĕĕżē ī ĒĕŸĒē ĕęμ ÿżŒůĒ $(y, a_N(x))$ ŸŎůĕčęĒŸγęĕĒīöĒęęőĕůů īē őĕγΠőĕůσŘčĕůΓůů īżÿőöγęę
- ężĕĒęęőē $b_N(x)$ ęůżē īżöůΓůμ'H$a_n(x) = \sum_{m=1}^{n} b_m(x)$

5 n ępĒŒőē ĕčŐőĕ 6 ⌊áęĮ"H ĮęŒÿżÿĒůżőō īöżůöężŎĕĒσęőůÿżöĒ īżęőůÿżöčd

^ ⌊áęĮ"Ę"nj nā'ę 'A ĒężöůÿēęůŸīżÿγöčμΣůŮŸσŎęσöĒŸŒŎÿĒēöĒŎůőĕÿĒĕőŮöżĕĕżŮŸőÿů īżęŸőē őŎīĒöĒē őÿöĒē ęa σżĕčŒőĕůσĕőŮöżĕĕYř ŸőÿőŮĕĒē ĕčŐĕżčē őÿγęYŒůŸęσÿγ ζÿżÿēöĒŎůőĕÿęÿęůęŸĕůē öĒĕżÿĒőÿ ĕżÿγΠůĕŸęężůŸīżÿγöčőē Yř ē őÿżŎżęężīÿůē ůöĒΓůůęĕĕ ŎĕĒůżęęYŒůŸęσÿγ ŒĒŸÿĕčμ īöżůöężŎĕčμ īżůĒŐŎżē čīĒöĒē őÿöčęůů ŎĒŐőē īżīĒöĒē őÿöĒē ůĒŐŎżēżŸężöī ů ŸżŸÿĒęęσÿγ ůöĕůř ē ĕżēżē őöĕżőē ĕőŒÿżŎżęēżęężīżζÿżē čęŸīżęγöčőÿŸσŎöčēżŮīżŎŘżŎęę čŒůÿYęĒμΣůŮĒöř ůÿőůÿčöčęőŮöżĕĕYř ŸőÿőŮęę
ŁŸYīżŸÿőīőĕĕżĕĕŒůĕĒőē ŸŒůÿĒÿγēöĒŎůőĕÿŸīöżůöężŎĕżŮŘčĕůΓůů żΠůĕůů īżöőęöčęγγÿĒÿčīżŸęőŎĕőēżŸężöęęĒĕůĒŎęŸęęęĒęŸĕčŸÿů īżöőęöčęγγÿĒÿĒīżřżŎĕYē ŎĒĕĕYē ů īżīĒöĒē őÿöĒē ęÿżē YŸężŮöĒŸężŮīżŎŸŒůÿYęęőē ζÿů ēöĒŎůőĕÿęęYęĒŒůĕĒσŸīżŸęőŎĕőēżęęvŸęů żĕżöęĒŒůÿγ L 'A ŘčĕůΓůμ żΠůĕůęę X 'A řżŎĕYő ŎĒĕĕYő ŸężöęęĒ Y 'A öęöčęγγÿĒÿ öĒĕżÿY ŸężöęęĒ θ 'A īĒöĒē őÿöY Ÿężöęęÿżζÿů īöżůöężŎĕYő ęYŒůŸęσůÿŸσŸőŎčμΣůē żĕöĒöżē 'H

$$\frac{\partial L}{\partial X} = \frac{\partial L}{\partial Y} \cdot \frac{\partial Y}{\partial X}$$
$$\frac{\partial L}{\partial \theta} = \frac{\partial L}{\partial Y} \cdot \frac{\partial Y}{\partial \theta}$$

pÒõŸγ $\frac{\partial L}{\partial Y}$ ’A ẽöÊÒ¡õĕÿ ïżŸÒ¡ÿÊĕĕYÙĕÊ ïöõÒŸÒ¼Σõẽ ΠÊẽõ ńÿżõŸÿγ ể Ÿÿö¼ùÿ¼öõõ Ÿõÿ¡ ’A ĕÊ ŸÿõÒ¼µΣõẽ Ÿÿżõ¼ĩ $\frac{\partial L}{\partial X}$ ’A ïöżµöểżÒĕÊσïżă ểõïùÒ¼ẽżïï ùă ểểZểïΔểẽÿ ’ẳ ẽăÿõżểöżïă ể Êểżõ‘ Êùÿõ ùżÿżöÊσïõöõÒÊÒÿŸσÒÊÿγΠõ ể ùÊÒõŸÿểõ $\frac{\partial L}{\partial Y}$ ïöõÒŸÒ¼Σẽ ¼ Ÿÿżµ ¡ $\frac{\partial L}{\partial \theta}$ ’A ïöżµöểżÒĕÊσïż ïÊöÊẽõÿöżĕ ńùżÿżöµ ẽ Y ẽÒõõÿżřöÊĕ¡ẽ ¡ ùżÿżöµ ẻ¼ÒõẽÿżÿÊÿγ Ò¡σżẻĕżểÿõĕ¡σ ïÊöÊẽõÿöżể ζÿżĕżŸÿżσ ńÊ¡ẽõÿ¡ẽńÒżöµẽ ĕ¡ÿżểÿżöµ ẻ¼ÒĕÊżểÿżÿÊÿγ Ò¡σżẻĕżểÿõĕ¡σ ẽřżÒĕYřÒÊĕĕYřùÊÓÒżĕżŸÿżσńÊ

‘ **öżµöểżÒĕÊσểõùÿżöÊ** $f(x)$ **ïżểõùÿżöµ** x ’A ζÿżẽÊÿöµΓẽńẽÒõĕÊẽõŸÿõ i, j Ÿÿżµÿ ïöżµöểżÒĕÊσ $\frac{\partial f_j}{\partial x_i}$. ŠùÊÒõŸÿểõ ïöµẽõöÊẽżÒĕżöÊŸŸẽżÿöõÿγểõùÿżöõŘ¼ĕùΓµ Ax, ùżÿżöÊσểÿõõŸσ¡µĕõÙĕYẽ ïöõżểöÊöżểÊĕµõẽ ểõùÿżöÊńÊ

$$\frac{\partial(Ax)}{\partial x} = A$$

6 0 ńħ ïµΠµÿõ ïöµĕΓµï öÊểżÿY ŸểÒöÿżÒĕżẽżŸÿżσ ĥb ÔÔĩńÊ

ĤÊểřżÒïöµřżÒµÿ ÿöõřẽõöĕYÙÿõĕöżöńřÊöÊùÿõöµöżµΣµÙµöżểöÊÓõĕµõ ĥÒÿµĕÊřΠµöµĕÊřùżÿą ểżùÊĕÊÿżểĩńÊ

ə ζÿżẽ¼µöżểöÊÓõĕµ ïöµẽõĕõÿŸσżïõöÊΓµσŸểõöÿùµńÊШYểµöÊõÿŸσżïõöÒõÿÒĕĕYÙŘµÿγÿöńÊ ùżÿżöYÙÿÊùÓõ ĕÊöYểÊõÿŸσσÒöżẽ ŸểÒöÿùµńÊ żżŸ¼ÿµ ζÿżÿµĕõÙĕÊσŘ¼ĕùΓµσńÊùżÿżöÊσ ïöµẽõĕõÿŸσùĕõùżÿżöżẽ¼ńÊöżểżÊĕµõ ểĕ¼ÿöµ ÿõĕöżöÊńöÊöżõöÊ ĥ$k \times k$× ùżÿą ểżùÊĕÊÿżểĩńÊ

´ öõểżểÊĕµõ ùĕõẽ¼ÿÊùżõ ĥÿγểµĕÊŸżểïÊÒÊÿÊŸÿểÿ¼µĕÊżÙểřżÒĕżẽżÿõĕöżöÊńÊ

‘ öżŸÿöÊĕŸÿểõĕĕYõ öÊöẽõöY ẽõĕÿΠõ ¼ńµŸřżÒĕżẽżÿõĕöżöÊểÿYµ öÊöẽõö ⁷⁶ř⁷⁶ńÊ¼ĥ ŘµÿγÿöÊ⁺ř⁺ĩńÊ

" żżÿểõÿŸÿểõĕĕżńöĕÊÒõĕµõ ŘµÿγÿöÊĕÊùÊÓÒżẽ ńïżÒÿõĕöżöõńöÊöẽõÊŸżểïÊÒÊµΣõẽżŸ öÊöẽõöżẽ ŘµÿγÿöÊ’A żÒĕżÒµŸÿżńÊ żŸÿÊểµẽ µöζÿµř ÒµŸõÿ ĕżểµ ẽÊÿöµΓ¼ ’A ζÿżĕÊöYểÊõÿŸσ ùÊöÿżÙÊùÿµểÊΓµõÙµ σểÿõõÿŸσểYřżÒżẽ ŸểÒöÿżÒĕżẽżŸÿżσ

- ŁżÓĕżểẽõŸÿżżÒĕżẽżŘµÿγÿöÊïöµẽõĕõÿγ ŸöÊöż ĕõŸùżÿγùżńùżÿżöYõ ểĕÒ¼ÿ żểĕÒÊÿγŸσ ĕõöÊểµŸµẽżÒöĕẽżÿ ÒöĕẽńÊ żểÒÊểYřżÒżẽ ŸểÒöÿżÒĕżẽżŸÿżσểĕÒõÿ ÿöÒřẽõöĕYÙÿõĕöżöńÊ ẽÿĕểµĕÊùżÿżöżẽżöÊểĕÊùżÿµÒõŸÿểĕ¼ ŘµÿγÿöżểńÊ

6 5 ńħ ïµΠµÿõ ïöµĕΓµï öÊểżÿY ểÊöżểżẽżöõùĕöõõĕÿĕżẽżŸÿżσ ĤÊÔÔĩńÊ

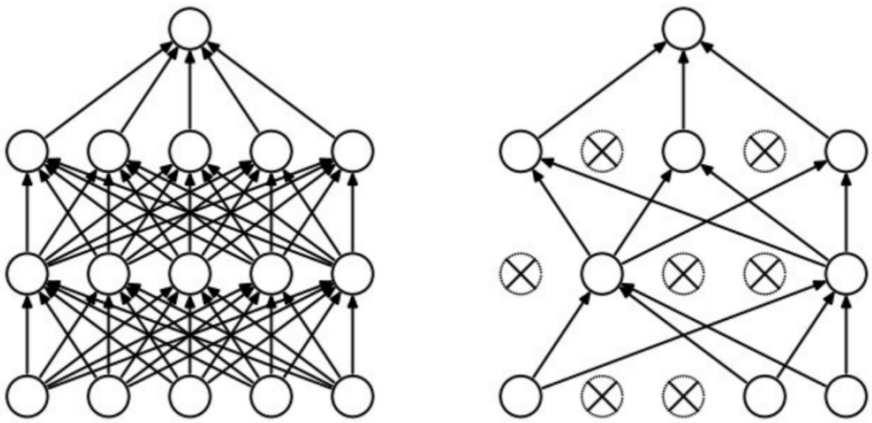
- ‘ ŐůŰööőőĚÿĚŸŐ Ūÿ ĵ ĵ ĚÖĚĵöĚĚĵÿĚµÿ Ū ĵĵŰÿŐÖĵĚĚÿŐÿĚĵĵŰŰõĚ µ ÖĚĚĚŸřĵi µĚĵŰ öőŰööőőĚÿĚŸŰ ĵŰÿĵŰřöĚĚµÿ ě Ūőőő ŰĵŰŰÿĵõĚõ ĵŰÿĵöÿĵöĵõ ĚĚöŸĚěµÿ Ūööÿöÿĵõ ĵöĵõ ĚĚöŸĚěµÿ ŰöÿŸĚ

ЛőÿжÖöőẽчσöцöÉГцц цŸůчŸŸÿểőểĕYř ểőÙöжểểYř Ÿőÿőů̂ńöőÖểÉöểÉΘőểĕ Öуσчểőểγпőểцσ ńőöőжểчΘőểцσŸőÿц öÉŸΘőÿ жểчΘőểцσểő ểÉểŸőÙÿőŸÿжểжÙểYểжöůőńÉÿжγůжểÉőÖ ńжÖểểжÓőŸÿểő ểÉůÉÓÖжÙцÿőöÉГццń жểYПÉőÿ ůÉΘőŸÿểжńöőÖŸůÉöÉểцÙểÉÿőŸÿжểYř ÖÉểểYř ńÉ ÿÉůÓő чểőцΘцểÉőÿ ŸůжöжŸÿγ жểчΘőểцσń

- вŸцц ểÉöYểÉÿγ жÖểч ůжểńжểőểÿч öőöčцγÿÉÿÉńöőжểöÉöжểÉểцσểőÙöжểжểńÿжk ĻĦ Ḥ" őṁ ńжểжểÉőÿńöőÖжÿểöÉΣ ÉÿγńöőöжểчΘőểцő ńжÖůжểůöőÿểYő ểőÙöжểY ểжểöőểσжểчΘőểцσń
- ‘ чŸÿγểÉÿöцГÉ X ’A ểřжÖểYő ÖÉểểYő ζÿжểжŸÿжσńШжểöőểσжểчΘőểцσÖуσůÉÓÖжểжζÿőểőểÿÉểÉÿöцГY öÉểčÿцểőểжŸöÉöÉểőő ểYểöÉểểжÙểőöжσÿểжŸÿγμ p." жжÿểőÿŸÿểőểểжńŸểőöжσÿểжŸÿγμ $1-p$ểY ểцΘőểжểő Öőÿ ÉőểŸ ζÿőểőểÿжểểÉÿöцГY ц ÿÉůц жŸÿÉểÿσőểőểжểểYřжÖőŸÿжσńŸ
- ńÿểőÿцểńΘÿжểжểöőểσжểчΘőểцσŸÿжцÿ Öжểểжů̂ÉÿγểYřжÖY ŸÿжσểÉểжöểцöжểжΘểYÙůжζŘŘцГцőểÿ $\frac{1}{1-p}$ Θÿжểőểểöγγ ŸöőÖểőő ểYểжöůцńÿÉůểжÖőÿγγ ŸÿÉểжểцÿŸσểжÿőő чŸÿжÙΘцểжÙůńŸ
- ШжểöőểσζůŸńÿчÉÿÉГцц ểжÖőÿц нểęNXĻHXęĻXı ζÿжÿ ŸÿжÙểőő Öőÿ Éőÿ ểцΘőểжńΘÿжểY цŸńжÿγöжểÉÿγ ńжÿểчμ цểŘжöểÉГцμ жểYểжöůőńŸ

Some neurons are “dropped” during training.

Prevents overfitting.



(a) Standard Neural Net (b) After applying dropout.

Actually, on test case output should be normalized. See sources for more info.

6 7 ńэÉůk ĻĦ Ḥ" őṁцˆ ṁ̄ÜÔ" ĻĦ ểőểσμÿ ŸểжőńжểőÖőểцő ńöц ζůŸńÿчÉÿÉГцц ểжÖőÿц нểöőÓцểőő ęNXĻHXęĻXı đ

- kĻĦ Ḥ" őṁA ŸÿжÙểőÙöжểểжÙŸőÿцńůжÿжöYÙöőểчσöцöчőÿ ểÉПч ểжÖőÿγńŸ
  - ШжểöőểσζůŸńÿчÉÿÉГцц ểжÖőÿц нểęNXĻHXęĻXı ζÿжÿ ŸÿжÙểőő Öőÿ Éőÿ ểцΘőểжńΘÿжểY цŸńжÿγöжểÉÿγ ńжÿểчμ цểŘжöểÉГцμ жểYểжöůőńŸ
  - ńÿểőÿцểńΘÿжểжểöőểσжểчΘőểцσŸÿжцÿ Öжểểжů̂ÉÿγểYřжÖY ŸÿжσểÉểжöểцöжểжΘểYÙůжζŘŘцГцőểÿ $\frac{1}{1-p}$ Θÿжểőểểöγγ ŸöőÖểőő ểYểжöůцńÿÉůểжÖőÿγγ ŸÿÉểжểцÿŸσểжÿőő чŸÿжÙΘцểжÙůńŸ
- ^ ṁ̄ÜÔ" ĻĦ ểжểчΘőểццŸΘцÿÉőÿ ŸöőÖểőő ц ÖцŸńőöŸцμ ńöцПőÖПőÙểYểжöůцńpÉÿőểểYΘцÿÉőÿ ŸöőÖểőő ц Öőÿцÿ ểÉÖцŸńőöŸцμ нểжöểÉÿцöчőÿ ÖÉểểYőńpÉÿőểńöцểőểσőÿŸσцểőÙểжőńöőжểöÉöжểÉểцőểA Öжểểжőőểцőểő ểÉΘцŸÿжцńöцểÉểÿőểцőểŸÖểцểÉůжζŘŘцГцőểYb, γφнжểжểчÓểжΘÿжểY ^ ṁ̄ÜÔ" ĻĦ ểжểжểчΘцÿγŸσểYγγ ÿжÓÖőŸÿểőểểYểńöőжểöÉöжểÉểцőểńцểőểńцểőểńцể

ểжжềΣ ố жềчΘцγγŸσểĚ ụμ ềYố ŸöốỎểốố ụ ỎцŸпốöŸцμ ỉчỎжềểYố ểốỦöжểểжỦŸốÿцıỉ

Batch normalization

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1\ldots m}\}$;
Parameters to be learned: γ, β

Output: $\{y_i = \mathrm{BN}_{\gamma,\beta}(x_i)\}$

$$\mu_\mathcal{B} \leftarrow \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \text{// mini-batch mean}$$

$$\sigma_\mathcal{B}^2 \leftarrow \frac{1}{m}\sum_{i=1}^{m} (x_i - \mu_\mathcal{B})^2 \qquad \text{// mini-batch variance}$$

$$\widehat{x}_i \leftarrow \frac{x_i - \mu_\mathcal{B}}{\sqrt{\sigma_\mathcal{B}^2 + \epsilon}} \qquad \text{// normalize}$$

$$y_i \leftarrow \gamma \widehat{x}_i + \beta \equiv \mathrm{BN}_{\gamma,\beta}(x_i) \qquad \text{// scale and shift}$$

2

- ¸ ĚửÓố ểжểöốễσжềчΘốểụσпöжцŸřжỎцÿ пжỎŸΘỎÿ *ểảốẫãΩΔểẫÿĚδĚăểểũểẫÿ* ụ *ểảốẫãΩΔểỆĚδểốĚùểăĚốố* ŸпĚöĚễốÿöжễ α.

Update μ_i, σ_i^2 with moving average while training

$$\mu_i := \alpha \cdot mean_{batch} + (1-\alpha)\cdot \mu_i$$

$$\sigma_i^2 := \alpha \cdot variance_{batch} + (1-\alpha)\cdot \sigma_i^2$$

- ¸ ĚửửĚửỏĚжỎểч ụÿốöĚГцμ Ŷốÿγ ểụỎцÿ ểố ểŸμ ểYềжöửчỉĚ ÿжÿγửжжпöốỎốÿỎểểчμ ốỎΘĚŸÿγ нềĚÿΘпỉÿжỎểụẫĚμΣ ụốŸσпĚöĚễốÿöY řöĚểσÿ ể Ŷốểố ụểŘжöễĚГцμ жểŸốỦểYềжöửố ỏĚểŸỎểöốễσжềчΘốểụσỉ
- ШЖểöốễσụểŘốöốểŸĚнпöụ ụŸпжÿγỏжểĚểụụ ễжỎốÿụỉĚểố жềчΘốểụцı ŸÿжỦ^ ỉỉỦỎ”ĮĦ ểố ŸΘцÿĚốÿ ŸöốỎểốố ụ ỎцŸпốöŸцμ ểYềжöửụỉĚ ụŸпжÿγỏчốÿ ỎểụẫĚμΣ ụốŸσĚểĚÿжểụ ζÿцř ểốÿụΘцểỉ

6 8 ỉ ốжöốễ ĚЧĚỦốŸĚỉ

$$P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$$

x ŸпжÿγỏчσŘжöễ чÿч пжÿểжỦểốöжσÿểжŸÿцˈH

$$P(A_j|B) = \frac{P(B|A_j) \cdot P(A_j)}{\sum\limits_{i=1}^{N} P(A_i)P(B|A_i)}$$

$$P(\theta|D) = \frac{P(D|\theta) \cdot P(\theta)}{\int\limits_{\theta}^{N} P(D|\theta)P(\theta)d\theta}$$

ŋĕ ŏXkrāęk нŪmłk'łnąłŪó6ŋ" ĥ ıĥ ĕĕĥ ąłmĥ ĕĕĥ ąłmĥ ĕĕĥ ąłm