

Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心



自步学习 (Self-paced Learning)

赵惟肖 硕士

2018年1月1日

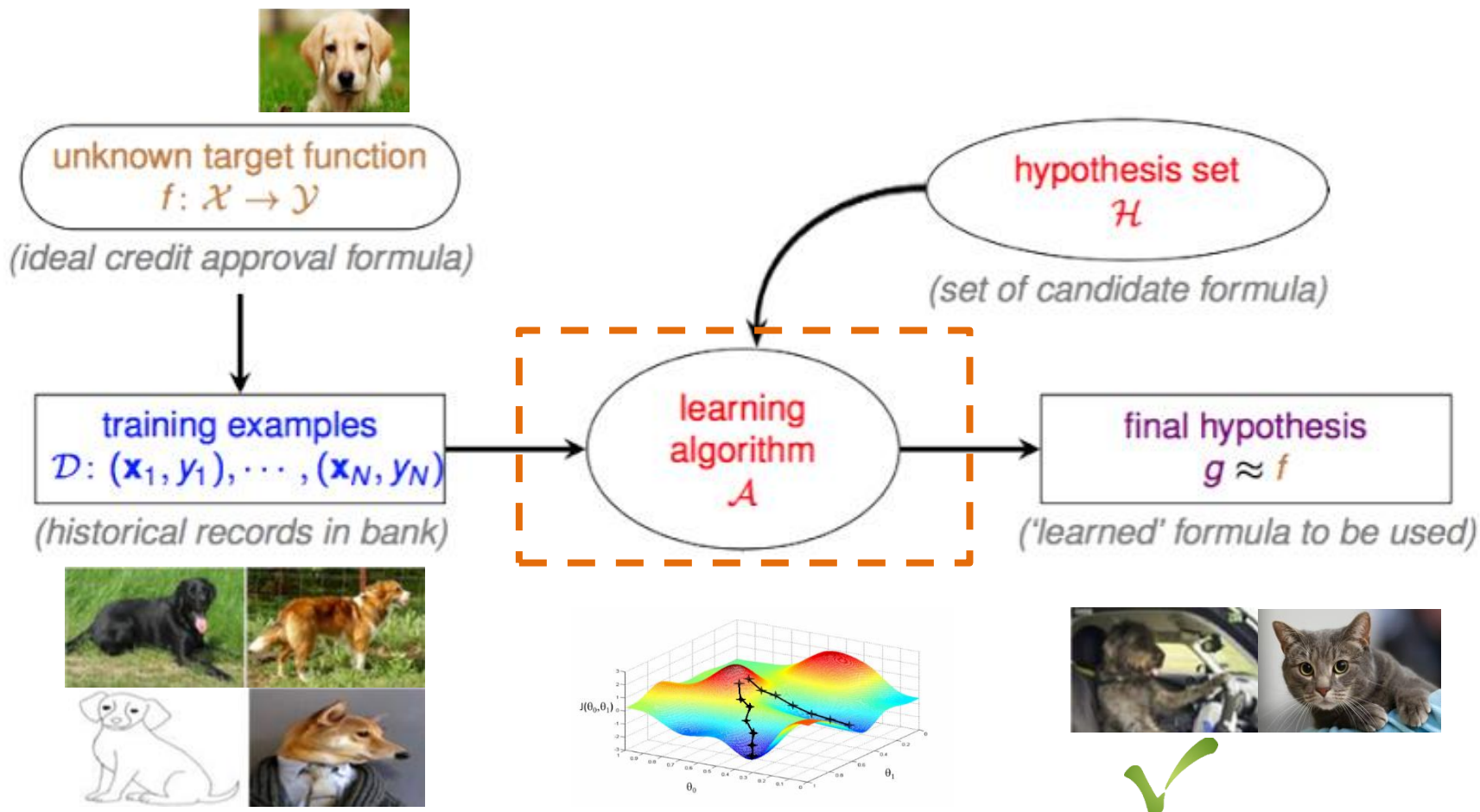


- 背景简介
- 基本概念
- 算法原理
- 延伸算法
- 应用总结
- 参考文献



背景简介

- 机器学习基础架构^[1]





基本概念

- 课程学习
 - 2009年Bengio教授在ICML上提出课程学习。
 - 通过模拟人的认知机理，首先学习简单的、普适性的知识结构，然后逐渐增加难度，过渡到学习更复杂、更专业化的知识。



Samples of “Dog” to learn earlier.



Samples of “Dog” to learn later.

- 课程学习
 - 2009年Bengio教授在ICML上提出课程学习。
 - 通过模拟人的认知机理，首先学习简单的、普适性的知识结构，然后逐渐增加难度，过渡到学习更复杂、更专业化的知识。



“bus” samples to learn earlier



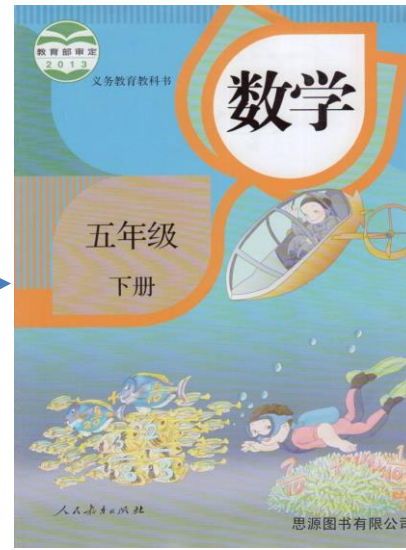
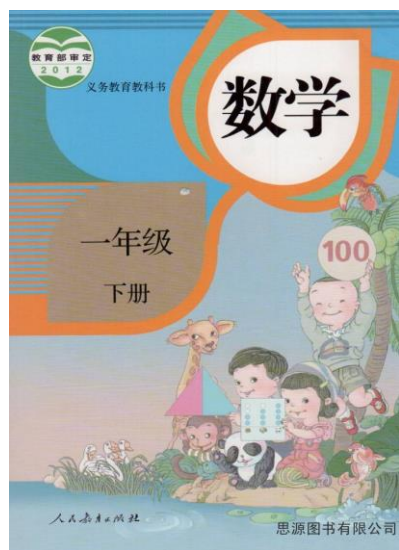
“bus” samples to learn later



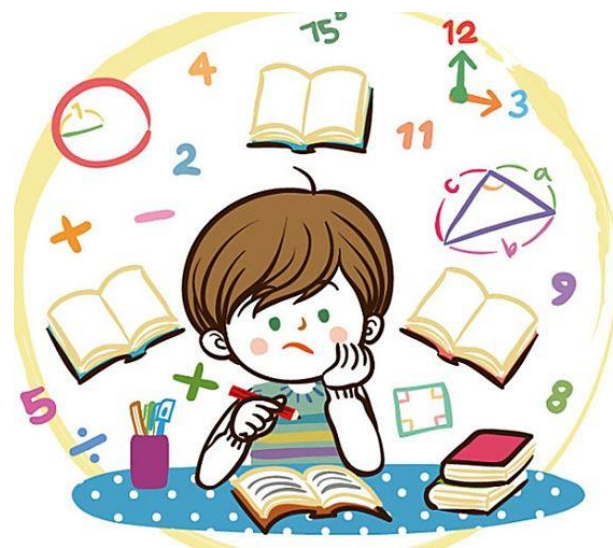
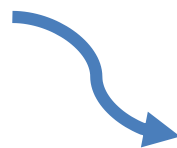
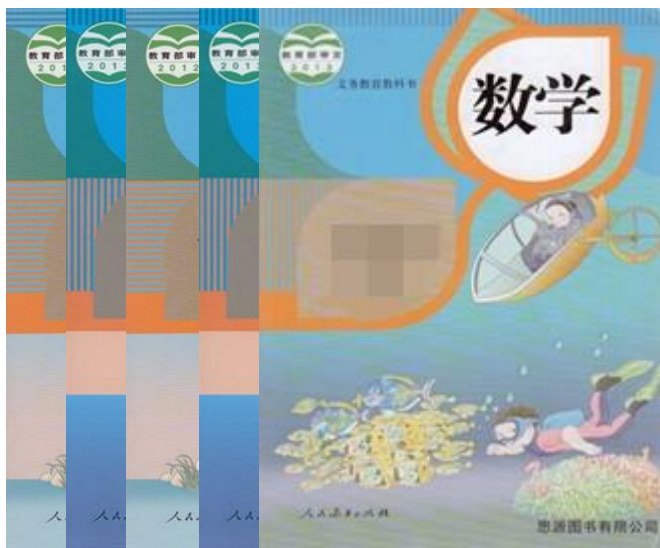
Age



- 自步学习在课程学习上的改进
 - 在课程学习的基础上，Koller教授在2010年NIPS上给出了自步学习的数学表达。
 - 课程学习：根据先验知识赋予样本学习先后顺序。



- 自步学习在课程学习上的改进
 - 在课程学习的基础上，Koller教授在2010年NIPS上给出了自步学习的数学表达。
 - 自步学习：学习算法在每一步迭代中决定下一步学习样本。

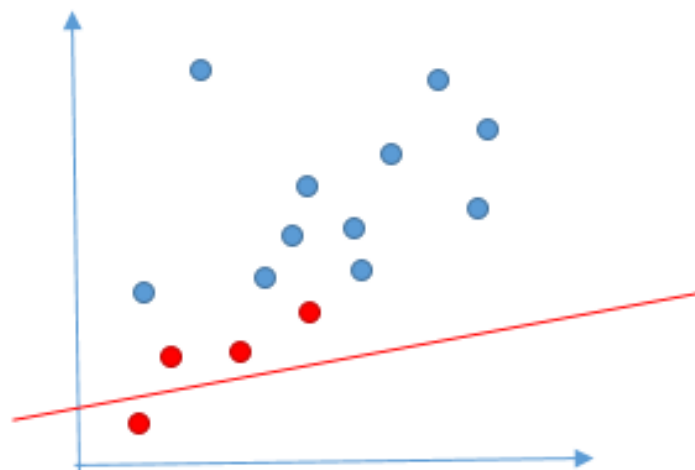
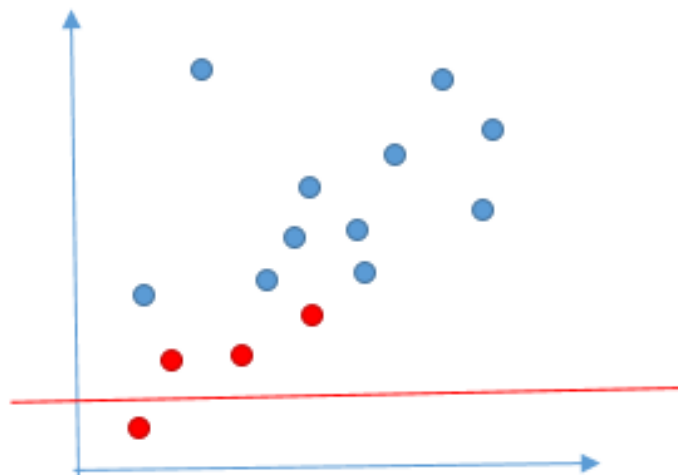
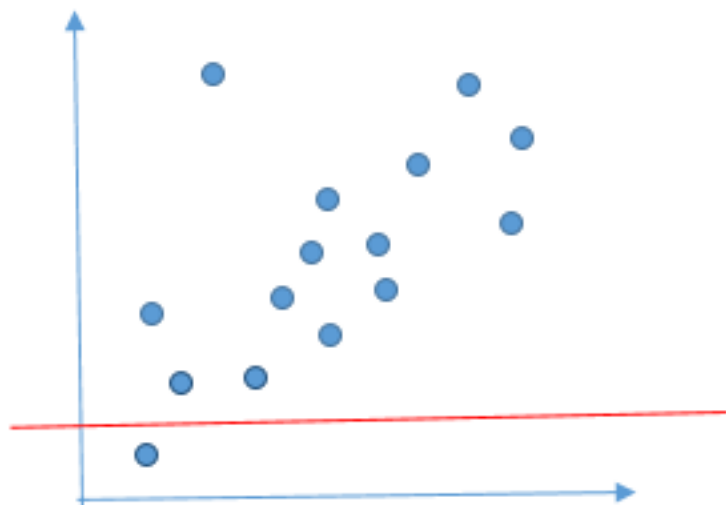




算法原理

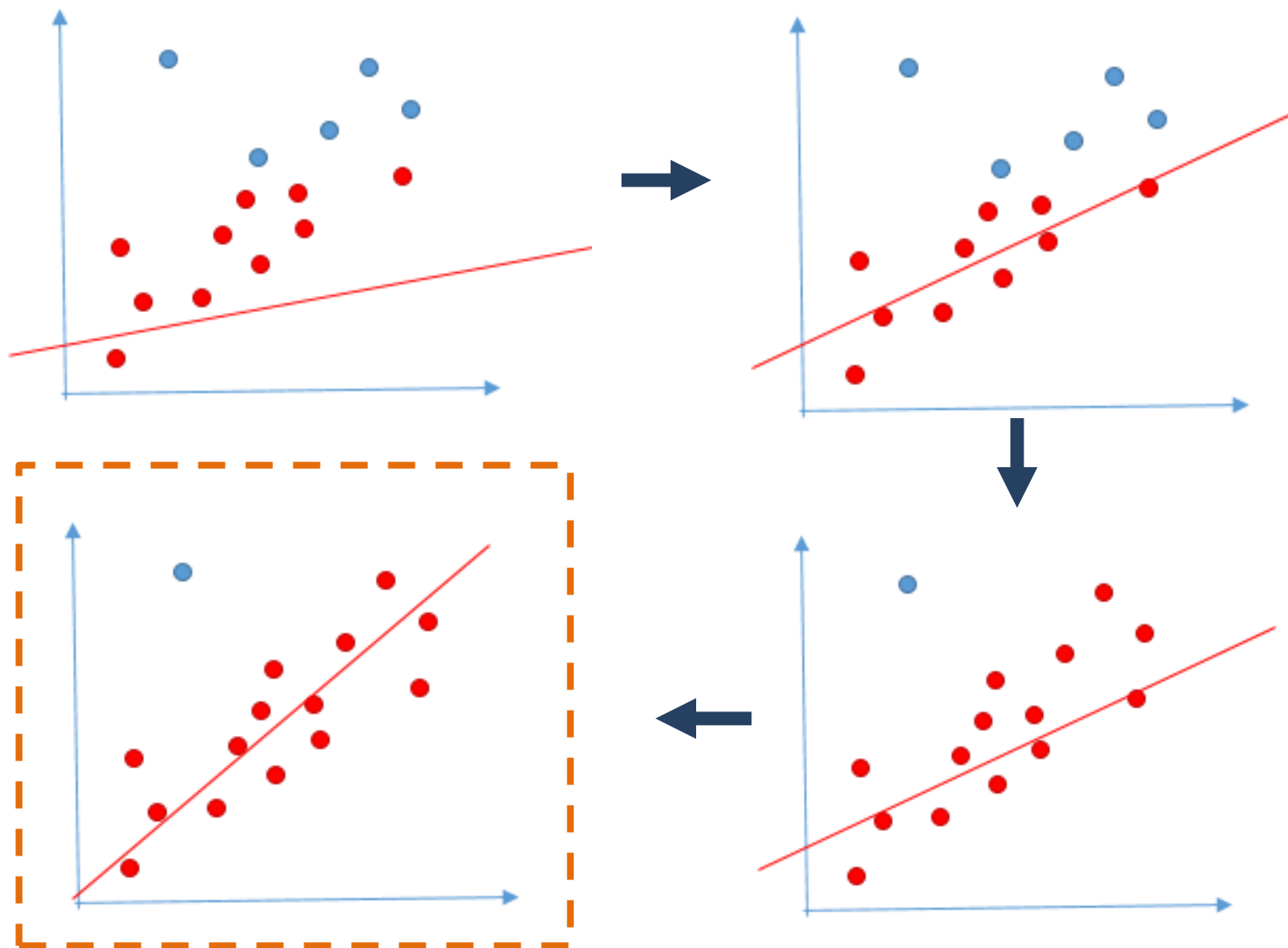


- 首先明确
 - 样本的选择在自步学习中并不是随机的，或是在一次迭代中全部纳入训练过程中，而是通过一种由简到难的有意义的方式进行选择的。
 - 简单样本可以理解为具有较小的损失（smaller loss）的样本。
 - 复杂的样本具有较大损失（larger loss）的样本。



● : 本次未被选择

● : 本次被选择





- 传统机器学习方法
 - 假设训练集 $D=(x_1, y_1), \dots, (x_n, y_n)$ 存在 n 个样本， x_i 代表第 i 个样本， y_i 表示对应样本的类别标签， w 表示对应模型的待优化参数，模型的损失函数记为 $f(x_i, y_i, w)$ 。传统机器学习方法的目标函数记为：

$$w_{t+1} = \operatorname{argmin} \left(\sum_{i=1}^n f(x_i, y_i; w) \right)$$

- 那么自步学习呢？



- 自步学习核心思想
 - 在每次迭代过程中倾向于从所有样本中选择具有很小的训练误差的样本，然后更新模型参数。
 - 因此，自步学习在传统机器学习目标函数中引入二分变量 v_i （1/0），用于表征每个样本是否被选择（是否为简单样本），其目标函数可以改为：

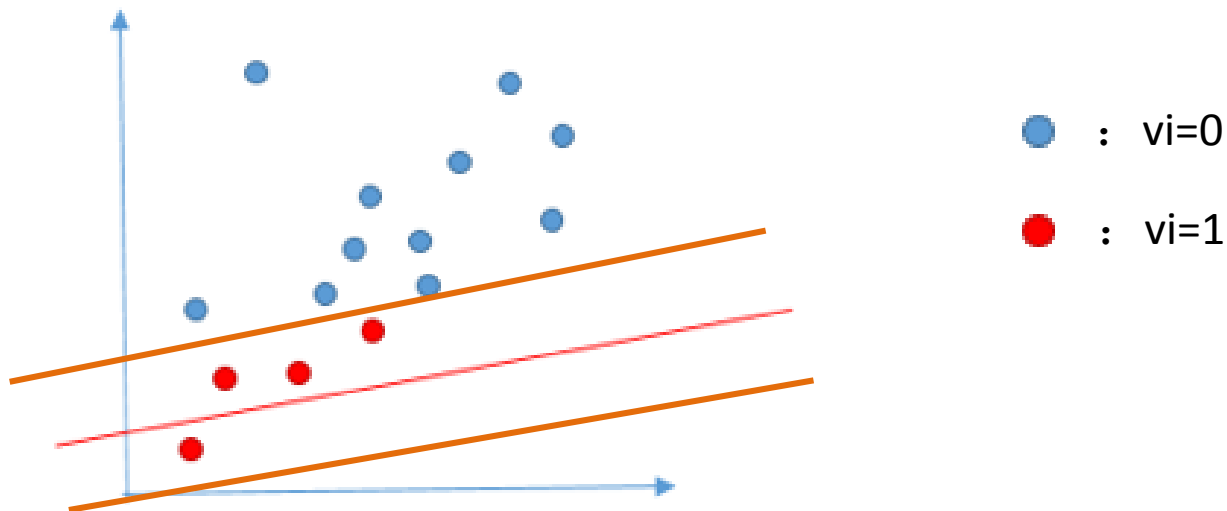
$$(w_{t+1}, v_{t+1}) = \operatorname{argmin} \left(\sum_{i=1}^n v_i f(x_i, y_i, w) \right)$$

出现问题： 如何选取 v_i 才能使损失最小→所有 v_i 都为0→不考虑任何损失→显然是有问题的。

- 更新了模型，使得在当 $f(x_i, y_i, w) \leq 1$ 时，损失项变得更小。

$$(w_{t+1}, v_{t+1}) = \operatorname{argmin} \left(\sum_{i=1}^n v_i f(x_i, y_i, w) - \sum_{i=1}^n v_i \right)$$
$$= \operatorname{argmin} \left(\sum_{i=1}^n v_i (f(x_i, y_i, w) - 1) \right)$$

出现问题：模型是无法调节的，每次都选择 $f \leq 1$ 时的样本





- 引入K，随着迭代次数的增加，我们逐渐减小K的值，当K值趋近于0时，越来越多的样本被选择。

$$(w_{t+1}, v_{t+1}) = \operatorname{argmin} \left(\sum_{i=1}^n v_i f(x_i, y_i, w) - \frac{1}{K} \sum_{i=1}^n v_i \right)$$

- 当 $f(x_i, y_i, w) \leq \frac{1}{K}$ 时， $v_i=1$ ，当 $f(x_i, y_i, w) > \frac{1}{K}$ 时， $v_i=0$ 。

- 自步学习的计算流程

$$\min E(\mathbf{w}, \mathbf{v}; \lambda) = \sum_{i=1}^n v_i L(x_i, y_i, \mathbf{w}) - \lambda \sum_{i=1}^n v_i$$

Algorithm 1: SPL Algorithm.

input : Input dataset \mathcal{D} , pace parameter $\mu > 1$, and self-paced function f .

output: Model parameter \mathbf{w}

```
1 Initialize  $\mathbf{w}^*, \lambda$ ; // assign the starting value
2 while not converged do
3   | while not converged do
4   |   | Update  $\mathbf{v}^* = \arg \min_{\mathbf{v}} \mathbb{E}(\mathbf{w}^*, \mathbf{v}; \lambda)$ ;
5   |   | Update  $\mathbf{w}^* = \arg \min_{\mathbf{w}} \mathbb{E}(\mathbf{w}, \mathbf{v}^*; \lambda)$ ;
6   |   end
7   |  $\lambda \leftarrow \mu \lambda$ ; // update the learning pace
8 end
9 return  $\mathbf{w} = \mathbf{w}^*$ ;
```



延伸算法

延伸算法



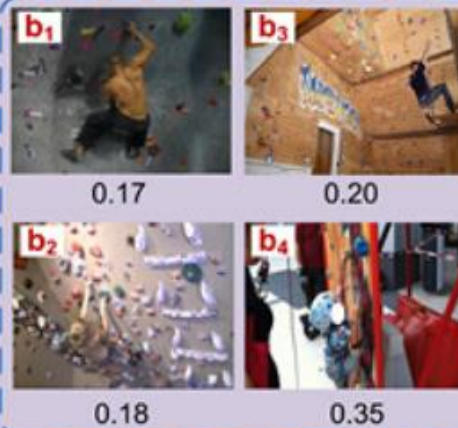
- Self-paced Learning with Diversity (SPLD)
SPLD formalizes the preference for both easy and diverse samples into a general regularizer.

Positive training samples of "Rock Climbing"

Outdoor bouldering



Artificial wall climbing



Snow mountain climbing



Curriculum for SPL



Curriculum for SPLD





- Self-paced Learning with Diversity (SPLD)

$$\min E(w, v; \lambda, \gamma) = \sum_{i=1}^n v_i L(x_i, y_i, w) - \lambda \sum_{i=1}^n v_i - \gamma \sum_{j=1}^b \|v^{(j)}\|_2$$

- b: 样本种类个数

- $\|v^{(j)}\|_2$: 第二范数 $\sqrt{\sum_{i=1}^n (v_i^j)^2}$

- 举例:

种类1: 10样本 种类2: 10样本

若全选择种类1: $\sqrt{10}$ 若各选择一半: $\sqrt{5} + \sqrt{5}$

$\sqrt{5} + \sqrt{5} > \sqrt{10}$ 多样性好则 $\sum_{j=1}^b \|v^{(j)}\|_2$ 大

- Self-paced Learning with Diversity (SPLD)

$$\min E(\mathbf{w}, \mathbf{v}; \lambda, \gamma) = \sum_{i=1}^n v_i L(x_i, y_i, \mathbf{w}) - \lambda \sum_{i=1}^n v_i - \gamma \sum_{j=1}^b \|\mathbf{v}^{(j)}\|_2$$

Algorithm 2: Algorithm of Self-Paced Learning with Diversity.

input : Input dataset \mathcal{D} , self-pace parameters μ_1, μ_2
output: Model parameter \mathbf{w}

```
1  if no prior clusters exist then cluster the training samples  $\mathbf{X}$  into  $b$  groups  $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(b)}$ ;  
2  Initialize  $\mathbf{v}^*, \lambda, \gamma$ ; // assign the starting value  
3  while not converged do  
4  |   Update  $\mathbf{w}^* = \arg \min_{\mathbf{w}} E(\mathbf{w}, \mathbf{v}^*; \lambda, \gamma)$ ; // train a classification model  
5  |   Update  $\mathbf{v}^* = \arg \min_{\mathbf{v}} E(\mathbf{w}^*, \mathbf{v}; \lambda, \gamma)$  using Algorithm 1; // select easy & diverse samples  
6  |    $\lambda \leftarrow \mu_1 \lambda; \gamma \leftarrow \mu_2 \gamma$ ; // update the learning pace  
7  end  
8  return  $\mathbf{w} = \mathbf{w}^*$ 
```

延伸算法



- Self-paced Learning with Diversity (SPLD)

Outdoor bouldering

Bear climbing a rock



0.28



0.05



0.12



0.15



0.12



0.12



0.40

Artificial wall climbing



0.17



0.35



0.18

Snow mountain climbing



0.15



0.20

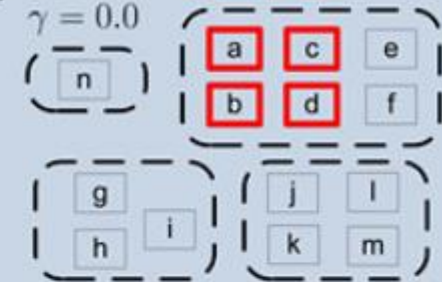


0.16

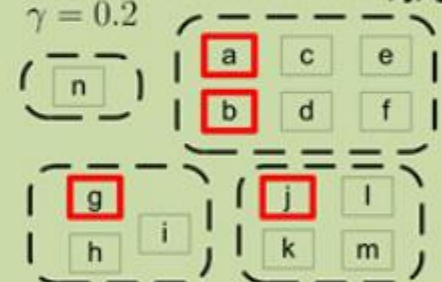


0.50

(a) $\lambda = 0.15$ Curriculum: a, b, c, d
 $\gamma = 0.0$



(b) $\lambda = 0.03$ Curriculum: a, j, g, b
 $\gamma = 0.2$



(c) $\lambda = 0.0$ Curriculum: a, j, g, n
 $\gamma = 0.285$



- Self-paced Curriculum Learning (SPCL)

CL



SPL



- Self-paced Curriculum Learning (SPCL)

$$\min E(\mathbf{w}, \mathbf{v}; \lambda, \Psi) = \sum_{i=1}^n v_i L(x_i, y_i, \mathbf{w}) - \lambda \sum_{i=1}^n v_i \quad \mathbf{v} \in \Psi$$

Algorithm 1: Self-paced Curriculum Learning.

input : Input dataset \mathcal{D} , predetermined curriculum γ , self-paced function f and a stepsize μ

output: Model parameter \mathbf{w}

- 1 Derive the curriculum region Ψ from γ ;
 - 2 Initialize \mathbf{v}^* , λ in the curriculum region;
 - 3 **while** *not converged* **do**
 - 4 Update $\mathbf{w}^* = \arg \min_{\mathbf{w}} \mathbb{E}(\mathbf{w}, \mathbf{v}^*; \lambda, \Psi)$;
 - 5 Update $\mathbf{v}^* = \arg \min_{\mathbf{v}} \mathbb{E}(\mathbf{w}^*, \mathbf{v}; \lambda, \Psi)$;
 - 6 **if** λ is small **then** increase λ by the stepsize μ ;
 - 7 **end**
 - 8 **return** \mathbf{w}^*
-



优劣分析



- Curriculum Learning **versus** Self-paced Learning

- CL:

优点：先验知识充分。

缺点：课程事先确定，可能与动态模型的最佳顺序不一致；人工标明顺序的难度和代价太大。

- SPL:

优点：无需事先确定学习顺序；且每次迭代中选择置信度高的样本，噪声和异常值由于置信度低而不被选择，有健壮（Robust）的学习机制。

缺点：缺乏先验知识。

Difficult to judge which one is better in practice.



应用总结



- 图片相关度重排序

L. Jiang, D. Meng, T. Mitamura, and A.G. Hauptmann. Easy samples first: Self-paced reranking for zero-example multimedia search. MM, 2014, 547 - 556.

- 非凸矩阵分解

Q. Zhao, D. Meng, L. Jiang, Q. Xie, Z. Xu, and A.G. Hauptmann. Self-Paced Learning for Matrix Factorization. AAAI, 2015, 3196 - 3202.

- 视频中长时间的物体跟踪

J.S. Supancic and D. Ramanan. Self-paced learning for long-term tracking. CVPR, 2013, 2379 - 2386.

-



参考文献



- [1] Bengio Y, Louradour J, Collobert R, et al. Curriculum learning[J]. Journal of the American Podiatry Association, 2009, 60(60):6.
- [2] Kumar M P, Packer B, Koller D. Self-paced learning for latent variable models. NIPS, 2010, 1189-1197.
- [3] L. Jiang, D. Meng, S.I. Yu, Z. Lan, S. Shan, and A.G. Hauptmann. Self-paced learning with diversity. NIPS, 2014, 2078–2086.
- [4] Lu Jiang, Deyu Meng, Qian Zhao, Shiguang Shan, Alexander Hauptmann. Self-paced Curriculum Learning. AAI, 2015
- [5] D. Zhang, D. Meng, C. Li, L. Jiang, Q. Zhao, and J. Han. A Self-paced Multiple-instance Learning Framework for Co-saliency Detection. In ICCV, 2015.



世界上本没有路 但只要努力自己 (Self) 去走一走 (Pace) 看上去似乎好像就有路了。

——鲁迅

知人者智，自知者明。
胜人者有力，自胜者
强。知足者富。强行
者有志。不失其所者
久。死而不亡者，寿。

谢谢！

