

在线学习及其应用

张利军

南京大学

计算机科学与技术系



目录

➤ 研究背景

- 机器学习

- 大数据

➤ 在线学习

- 定义

- 完全信息在线学习

- 赌博机在线学习

➤ 总结与展望

目录

➤ 研究背景

- 机器学习

- 大数据

➤ 在线学习

- 定义

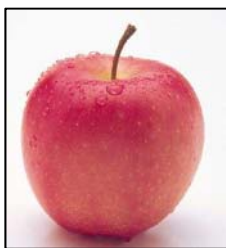
- 完全信息在线学习

- 赌博机在线学习

➤ 总结与展望

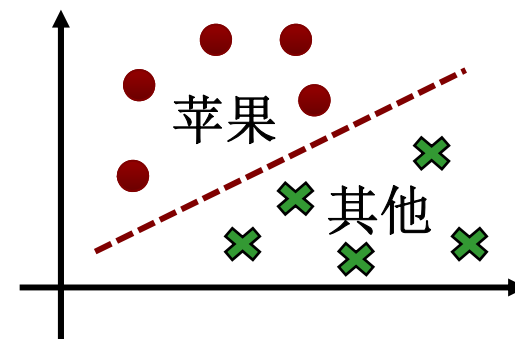
机器学习

➤ 监督学习



苹果

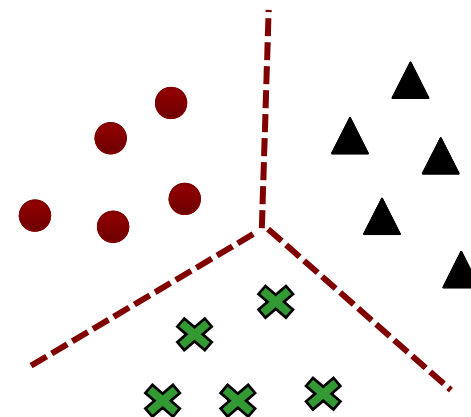
$$\begin{matrix} (\mathbf{x}_1, y_1) \\ \dots \\ (\mathbf{x}_n, y_n) \end{matrix} \Rightarrow y \approx h(\mathbf{x})$$



➤ 无监督学习



$$\begin{matrix} \mathbf{x}_1 \\ \dots \\ \mathbf{x}_n \end{matrix} \Rightarrow h(\mathbf{x})$$



➤ 算法输入

□ 训练数据: $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n) \in \mathcal{X} \times \mathcal{Y}$

□ 函数空间: $\mathcal{H} = \{h: \mathcal{X} \mapsto \mathbb{R}\}$

➤ 算法输出

□ $h \in \mathcal{H}$, 使得 $y_i \approx h(\mathbf{x}_i)$

➤ 经验风险最小化

$$\min_{h \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n \ell(h(\mathbf{x}_i), y_i)$$

□ $\ell(\cdot, \cdot)$ 代表某损失函数

典型学习算法(1)

➤ 最小二乘回归

$$\min_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^n (\mathbf{w}^\top \mathbf{x}_i - y_i)^2$$

□ 平方损失: $\ell(u, v) = (u - v)^2$

□ 线性函数空间: $\mathcal{H} = \{h(\mathbf{x}) = \mathbf{w}^\top \mathbf{x}\}$

➤ 线性支持向量机

$$\min_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i \mathbf{w}^\top \mathbf{x}_i) + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

□ 铰链损失: $\ell(u, v) = \max(0, 1 - uv)$

□ 线性函数空间: $\mathcal{H} = \{h(\mathbf{x}) = \mathbf{w}^\top \mathbf{x}\}$

□ 正则化: $\lambda \|\mathbf{w}\|^2 / 2$

典型学习算法 (2)

➤ 核支持向量机

$$\min_{h \in \mathcal{H}_\kappa} \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i h(x_i)) + \frac{\lambda}{2} \|h\|_{\mathcal{H}_\kappa}^2$$

□ 铰链损失: $\ell(u, v) = \max(0, 1 - uv)$

□ 核函数: $\kappa(\cdot, \cdot): \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$

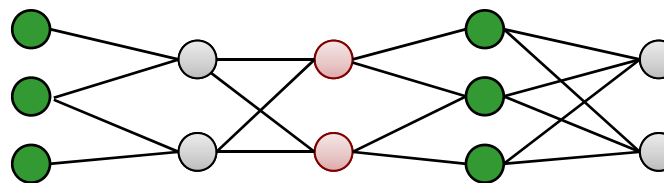
✓ 高斯核函数 $\kappa(\mathbf{x}, \mathbf{x}') = e^{-\|\mathbf{x} - \mathbf{x}'\|^2 / 2}$

□ 再生核希尔伯特空间 (RKHS): $\mathcal{H}_\kappa$

□ 正则化: $\lambda \|h\|_{\mathcal{H}_\kappa}^2 / 2$

➤ 深度学习

□ 函数空间: 神经网络



目录

➤ 研究背景

- 机器学习

- 大数据

➤ 在线学习

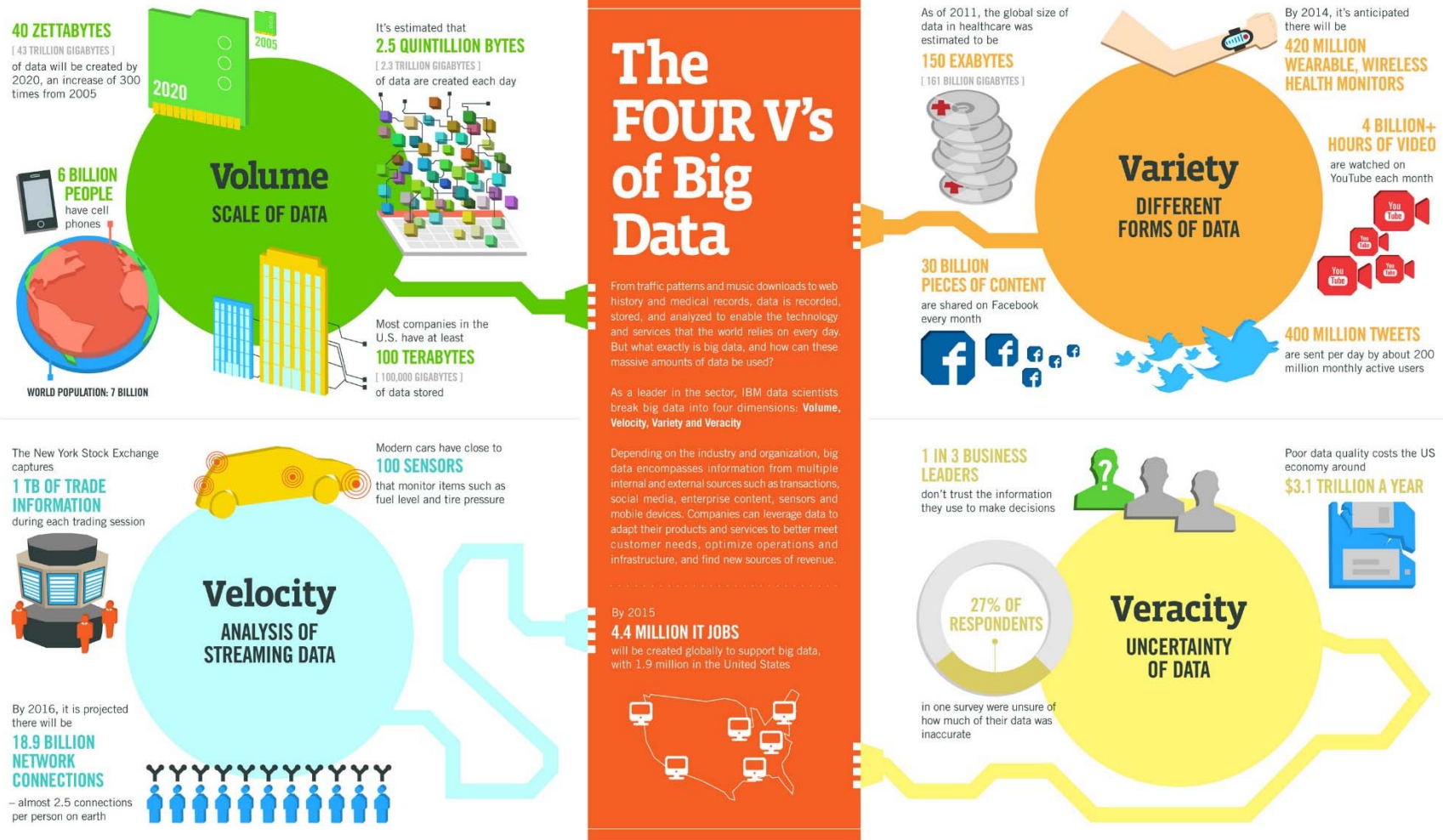
- 定义

- 完全信息在线学习

- 赌博机在线学习

➤ 总结与展望

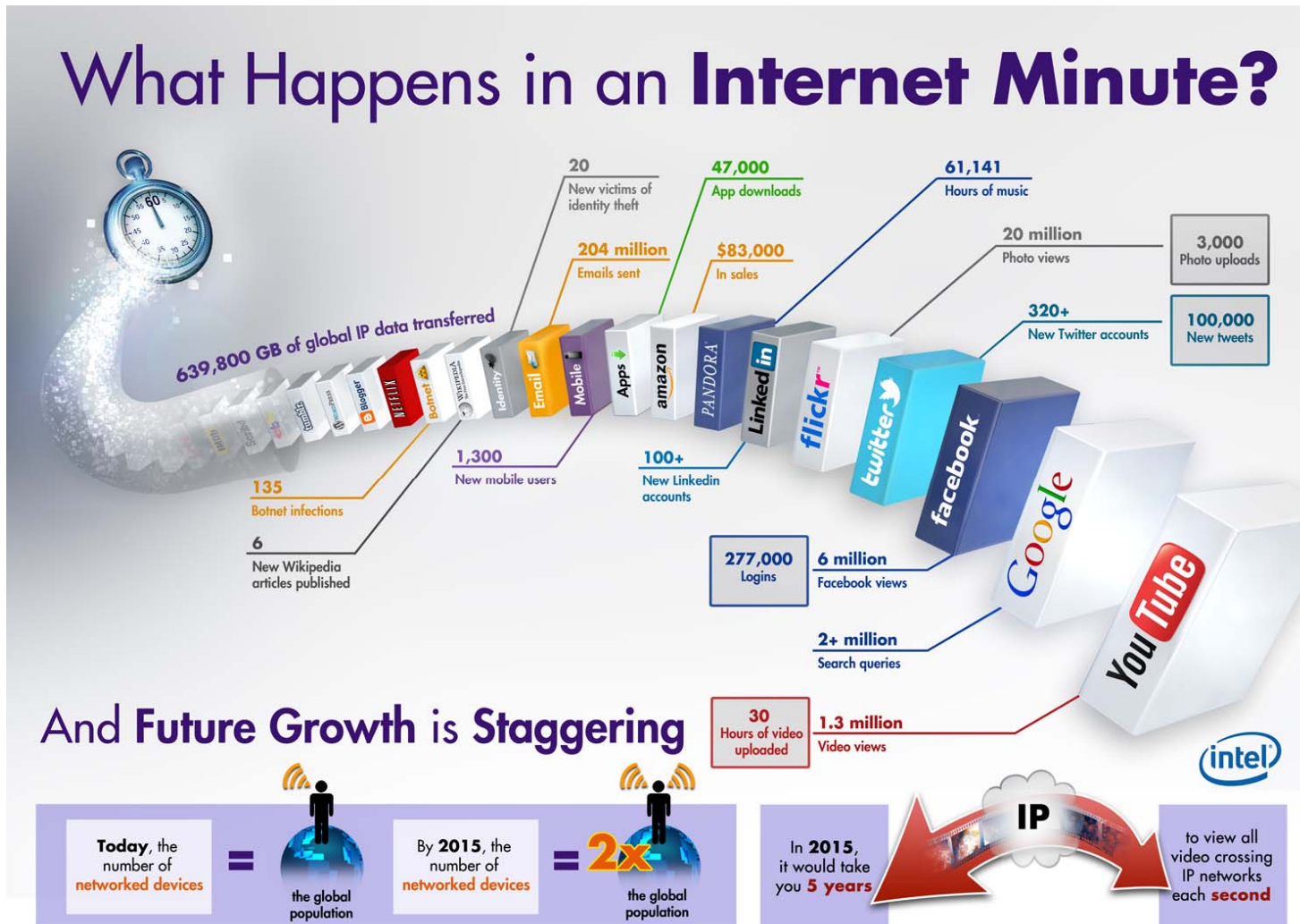
大数据



Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS

IBM

快速增长



大数据带来的挑战

➤ 经验误差最小化

$$\min_{h \in \mathcal{H}} \frac{1}{n} \sum_{i=1}^n \ell(h(\mathbf{x}_i), y_i)$$

□ n 是训练样本的数量

□ 训练复杂度: $O(n)$ 、 $O(n^2)$

➤ 大数据“规模大”

□ 训练过程计算量大

➤ 大数据“增长快”

□ 频繁需要重新训练

在线学习

目录

➤ 研究背景

- 机器学习

- 大数据

➤ 在线学习

- 定义

- 完全信息在线学习

- 赌博机在线学习

➤ 总结与展望

在线学习

➤ 定义 [Shalev-Shwartz, 2011]

Online learning is the process of answering a sequence of questions given (**maybe partial**) knowledge of answers to previous questions and possibly additional information.

□ 完全信息在线学习



在线学习

➤ 定义 [Shalev-Shwartz, 2011]

Online learning is the process of answering a sequence of questions given (**maybe partial**) knowledge of answers to previous questions and possibly additional information.

□ 赌博机在线学习(Bandits)



数学定义

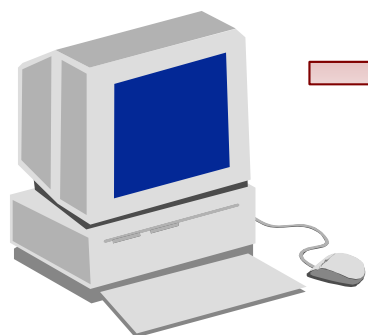
➤ 在线学习

for $t = 1, 2, \dots, T$ **do**

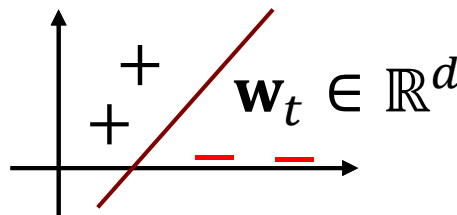
1. 学习器选择决策 $\mathbf{w}_t \in \mathcal{W}$

同时，对手选择函数 $f_t(\cdot)$

end for



分类器



样本 $(\mathbf{x}_t, y_t) \in \mathbb{R}^d \times \{\pm 1\}$

损失 $f_t(\mathbf{w}) = \max(1 - y_t \mathbf{w}^\top \mathbf{x}_t, 0)$



数学定义

➤ 在线学习

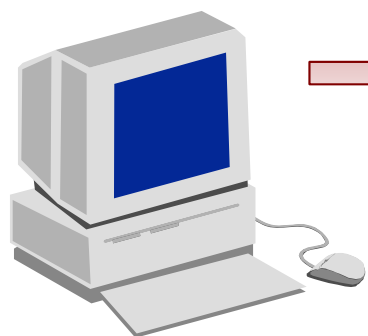
for $t = 1, 2, \dots, T$ **do**

1. 学习器选择决策 $\mathbf{w}_t \in \mathcal{W}$

同时，对手选择函数 $f_t(\cdot)$

2. 学习器遭受损失 $f_t(\mathbf{w}_t)$ ，并更新 $\mathbf{w}_t$

end for



遭受损失 $f_t(\mathbf{w}_t)$ ，更新 $\mathbf{w}_t$



数学定义

➤ 在线学习

for $t = 1, 2, \dots, T$ **do**

1. 学习器选择决策 $\mathbf{w}_t \in \mathcal{W}$

同时，对手选择函数 $f_t(\cdot)$

2. 学习器遭受损失 $f_t(\mathbf{w}_t)$ ，并更新 $\mathbf{w}_t$

end for

➤ 累积损失

$$\sum_{t=1}^T f_t(\mathbf{w}_t)$$

➤ 最小化累积损失

数学定义

➤ 在线学习

for $t = 1, 2, \dots, T$ **do**

1. 学习器选择决策 $\mathbf{w}_t \in \mathcal{W}$

同时，对手选择函数 $f_t(\cdot)$

2. 学习器遭受损失 $f_t(\mathbf{w}_t)$ ，并更新 $\mathbf{w}_t$

end for

➤ 遗憾 (Regret)

$$\sum_{t=1}^T f_t(\mathbf{w}_t) - \min_{\mathbf{w} \in \mathcal{W}} \sum_{t=1}^T f_t(\mathbf{w})$$

离线最优

➤ 最小化遗憾

数学定义

➤ 在线学习

for $t = 1, 2, \dots, T$ **do**

1. 学习器选择决策 $\mathbf{w}_t \in \mathcal{W}$

同时，对手选择函数 $f_t(\cdot)$

2. 学习器遭受损失 $f_t(\mathbf{w}_t)$ ，并更新 $\mathbf{w}_t$

end for

➤ 遗憾 (Regret)

$$\sum_{t=1}^T f_t(\mathbf{w}_t) - \min_{\mathbf{w} \in \mathcal{W}} \sum_{t=1}^T f_t(\mathbf{w}) = o(T)$$

➤ 最小化遗憾

目录

➤ 研究背景

- 机器学习
- 大数据

➤ 在线学习

- 定义
- 完全信息在线学习
- 赌博机在线学习

➤ 总结与展望

完全信息在线学习 (1)

➤ 感知机 (Perceptron) [Rosenblatt, 1958]

□ 分类算法

for $t = 1, 2, \dots, T$ **do**

1. 收到训练样本 $(\mathbf{x}_t, y_t)$

2. 如果 $y_t(\mathbf{w}_t^\top \mathbf{x}_t) \geq 0$

$$\mathbf{w}_{t+1} = \mathbf{w}_t$$

3. 否则

$$\mathbf{w}_{t+1} = \mathbf{w}_t + y_t \mathbf{x}_t$$

end for

□ 数据线性可分时算法收敛

完全信息在线学习 (2)

➤ 在线梯度下降 [Zinkevich, 2003]

□ 在线凸优化

for $t = 1, 2, \dots, T$ **do**

1. 看到损失函数 $f_t(\cdot)$
2. 遭受损失 $f_t(\mathbf{w}_t)$
3. 更新 $\mathbf{w}_t$

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \nabla f_t(\mathbf{w}_t)$$

end for

说明: $\eta_t \geq 0$ 是步长, $\nabla f_t(\mathbf{w}_t)$ 为 f_t 在 $\mathbf{w}_t$ 处的梯度

完全信息在线学习 (2)

➤ 在线梯度下降 [Zinkevich, 2003]

□ 关键步骤

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \nabla f_t(\mathbf{w}_t)$$

□ 理论保障

$$\sum_{t=1}^T f_t(\mathbf{w}_t) - \min_{\mathbf{w} \in \mathcal{W}} \sum_{t=1}^T f_t(\mathbf{w}) = \begin{cases} o(\sqrt{T}), & \text{凸函数} \\ o(\log T), & \text{强凸函数} \end{cases}$$

✓ 次线性、最优

在线支持向量机

➤ 线性支持向量机

$$\min_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i \mathbf{w}^\top \mathbf{x}_i) + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

□ 复杂度: $O(dn)$

➤ 在线版本——Pegasos [Shalev-Shwartz et al., 2007]

□ 给定样本 $(\mathbf{x}_t, y_t)$

□ 定义 $f_t(\mathbf{w}) = \max(0, 1 - y_t \mathbf{w}^\top \mathbf{x}_t) + \frac{\lambda}{2} \|\mathbf{w}\|^2$

□ 更新算法

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \nabla f_t(\mathbf{w}_t)$$

□ 复杂度 $O(d)$, 遗憾上界 $O(\log T)$

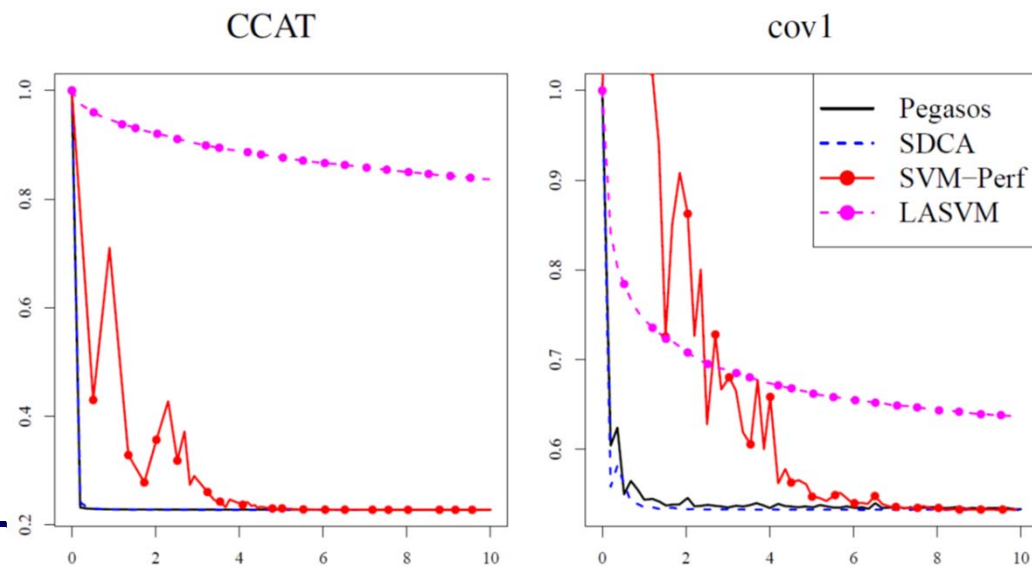
在线支持向量机

➤ 线性支持向量机

$$\min_{\mathbf{w}} \frac{1}{n} \sum_{i=1}^n \max(0, 1 - y_i \mathbf{w}^\top \mathbf{x}_i) + \frac{\lambda}{2} \|\mathbf{w}\|^2$$

□ 复杂度: $O(dn)$

➤ 在线版本——Pegasos [Shalev-Shwartz et al., 2007]



在线核学习

- 非线性支持向量机
- 非线性逻辑回归

$$\min_{h \in \mathcal{H}_K} \frac{1}{n} \sum_{i=1}^n \ell(h(\mathbf{x}_t), y_t)$$

- 在线核学习

- 给定样本 $(\mathbf{x}_t, y_t)$
- 定义 $f_t(h) = \ell(h(\mathbf{x}_t), y_t)$
- 更新算法

$$\mathbf{h}_{t+1} = \mathbf{h}_t - \eta_t \nabla f_t(\mathbf{h}_t)$$

- $\mathbf{h}_t$ 的存储复杂度 $O(t)$, 导致计算复杂度 $O(t)$

稀疏在线核学习

➤ 控制非线性函数的复杂度

□ [Zhang et al., AAAI 2012]

□ [Zhang et al., ICML 2013]

➤ 随机更新

□ 给定样本 $(\mathbf{x}_t, y_t)$

□ 定义 $f_t(h) = \ell(h(\mathbf{x}_t), y_t)$

□ 计算概率 p_t 来反映样本重要程度

□ 从伯努利分布采样 Z_t

□ 如果 $Z_t = 1$

$$\mathbf{h}_{t+1} = \mathbf{h}_t - \eta_t \nabla f_t(\mathbf{h}_t)$$

1. 设计概率
2. 理论分析

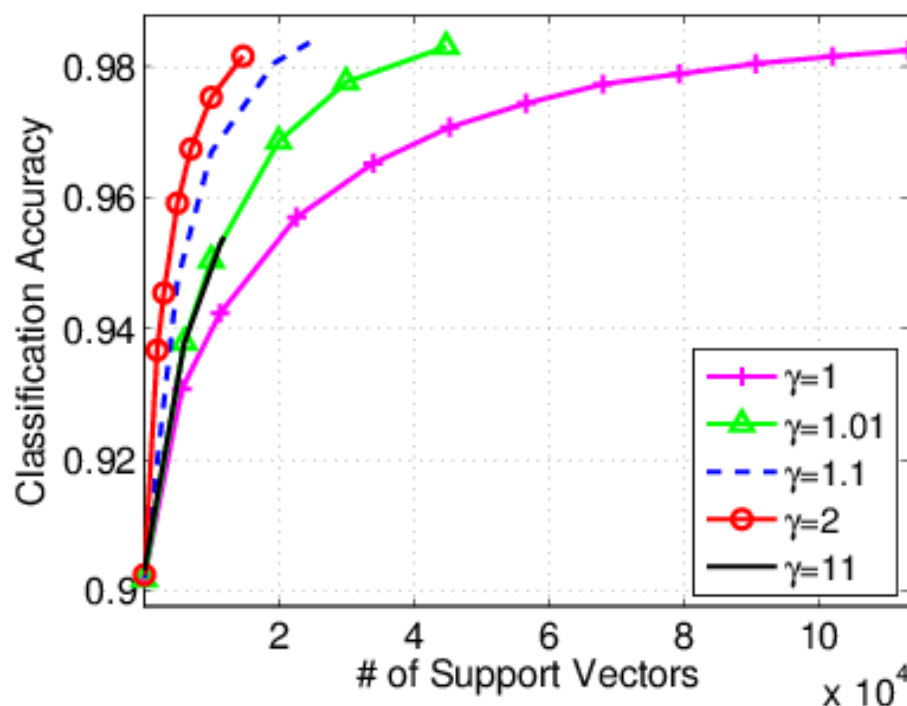
稀疏在线核学习

➤ 控制非线性函数的复杂度

□ [Zhang et al., AAAI 2012]

□ [Zhang et al., ICML 2013]

➤ 随机更新



目录

➤ 研究背景

- 机器学习

- 大数据

➤ 在线学习

- 定义

- 完全信息在线学习

- 赌博机在线学习

➤ 总结与展望

赌博机在线学习

- 多臂赌博机 [Robbins, 1952]
 - Multi-Armed Bandits (MAB)
 - 有 K 个臂，每次只能拉1个

手臂1	x_1^1
手臂2	x_1^2
手臂3	x_1^3



赌博机在线学习

- 多臂赌博机 [Robbins, 1952]
 - Multi-Armed Bandits (MAB)
 - 有 K 个臂，每次只能拉1个

手臂1	x_1^1
手臂2	10
手臂3	x_1^3



赌博机在线学习

- 多臂赌博机 [Robbins, 1952]
 - Multi-Armed Bandits (MAB)
 - 有 K 个臂，每次只能拉1个

手臂1	X_1^1	X_2^1
手臂2	10	X_2^2
手臂3	X_1^3	X_2^3



赌博机在线学习

- 多臂赌博机 [Robbins, 1952]
 - Multi-Armed Bandits (MAB)
 - 有 K 个臂，每次只能拉1个

手臂1	X_1^1	X_2^1
手臂2	10	X_2^2
手臂3	X_1^3	0



赌博机在线学习

- 多臂赌博机 [Robbins, 1952]
 - Multi-Armed Bandits (MAB)
 - 有 K 个臂，每次只能拉1个

手臂1	X_1^1	X_2^1	X_3^1
手臂2	10	X_2^2	X_3^2
手臂3	X_1^3	0	X_3^3



赌博机在线学习

- 多臂赌博机 [Robbins, 1952]
 - Multi-Armed Bandits (MAB)
 - 有 K 个臂，每次只能拉1个

手臂1	X_1^1	X_2^1	6
手臂2	10	X_2^2	X_3^2
手臂3	X_1^3	0	X_3^3



赌博机在线学习

- 多臂赌博机 [Robbins, 1952]
 - Multi-Armed Bandits (MAB)
 - 有 K 个臂，每次只能拉1个

手臂1	X_1^1	X_2^1	6	X_4^1
手臂2	10	X_2^2	X_3^2	X_4^2
手臂3	X_1^3	0	X_3^3	X_4^3



赌博机在线学习

- 多臂赌博机 [Robbins, 1952]
 - Multi-Armed Bandits (MAB)
 - 有 K 个臂，每次只能拉1个

手臂1	X_1^1	X_2^1	6	X_4^1
手臂2	10	X_2^2	X_3^2	0
手臂3	X_1^3	0	X_3^3	X_4^3



赌博机在线学习

➤ 多臂赌博机 [Robbins, 1952]

❑ Multi-Armed Bandits (MAB)

❑ 有 K 个臂，每次只能拉1个

手臂1	X_1^1	X_2^1	6	X_4^1	X_5^1
手臂2	10	X_2^2	X_3^2	0	X_5^2
手臂3	X_1^3	0	X_3^3	X_4^3	X_5^3

❑ 探索—利用困境

✓ Exploration vs Exploitation

➤ 置信上界法 (UCB) [Auer et al., 2002]

❑ 随机假设——独立同分布

❑ 遗憾界: $O(K \log T)$



置信上界法

- 假设有3个手臂
- 真实的均值（未知）为 μ_1 、 μ_2 、 μ_3

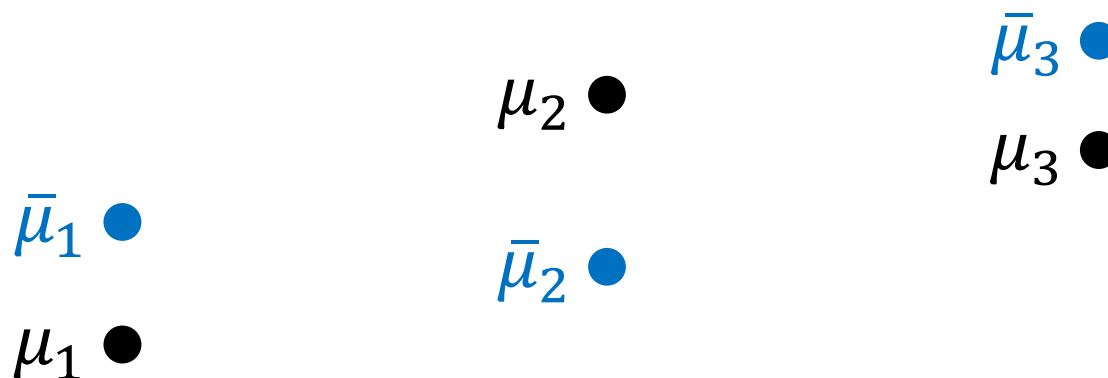
μ_1 ●

μ_2 ●

μ_3 ●

置信上界法

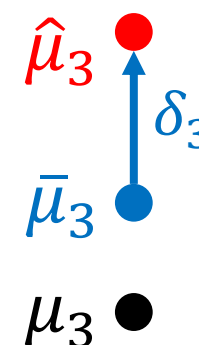
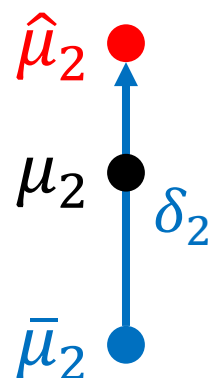
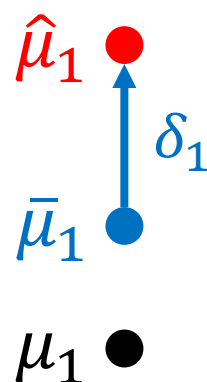
- 假设有3个手臂
- 真实的均值（未知）为 μ_1 、 μ_2 、 μ_3
- 估计值（样本平均）为 $\bar{\mu}_1$ 、 $\bar{\mu}_2$ 、 $\bar{\mu}_3$



置信上界法

- 假设有3个手臂
- 真实的均值（未知）为 μ_1 、 μ_2 、 μ_3
- 估计值（样本平均）为 $\bar{\mu}_1$ 、 $\bar{\mu}_2$ 、 $\bar{\mu}_3$
- 构造置信上界 $\hat{\mu}_1$ 、 $\hat{\mu}_2$ 、 $\hat{\mu}_3$

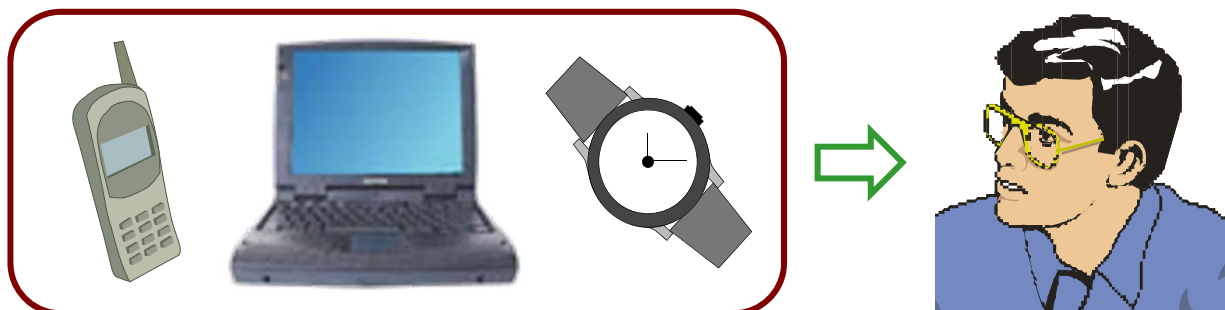
$$\square \mu_i \leq \hat{\mu}_i = \bar{\mu}_i + \delta_i$$



在线推荐系统

➤ 推荐系统

□ 学习器向用户推荐商品 $\mathbf{x}_t$



✓ 赌博机一次摇臂

□ 用户对商品打分 y_t

➤ 下一轮推荐的商品 $\mathbf{x}_{t+1}$?

□ 在线梯度下降 $\mathbf{x}_{t+1} = \mathbf{x}_t - \eta_t \frac{\partial y_t}{\partial \mathbf{x}_t}$

□ 用户的评分机制未知



随机线性赌博机

➤ 模型假设

- 学习器选择行为 $\mathbf{x}_t$
- 学习器观测反馈 $y_t \in \mathbb{R}$

局限于
实数反馈

$$y_t = \mathbf{x}_t^T \mathbf{w}_* + \epsilon_t$$

✓ $\mathbf{w}_*$ 是未知的参数, ϵ_t 为均值为0的随机噪声

➤ 核心问题

- 探索: 选择行为 $\mathbf{x}_t$ 来估计 $\mathbf{w}_*$
- 利用: 选择行为 $\mathbf{x}_t$ 来最大化 y_t

➤ 置信上界法 (UCB) [Dani et al., 2008]

- 遗憾界: $O(d\sqrt{T})$

基于二元反馈的线性赌博机

- 支持更加常见的二元反馈
 - 是否购买、是否点击、是否喜欢
 - [Zhang et al., ICML 2016]

- 模型假设

- 学习器选择行为 $\mathbf{x}_t$
 - 学习器观测反馈 $y_t \in \{\pm 1\}$

$$\Pr[y_t = \pm 1 | \mathbf{x}_t] = \frac{1}{1 + \exp(-y_t \mathbf{x}_t^\top \mathbf{w}_*)}$$

✓ $\mathbf{w}_*$ 是未知的参数

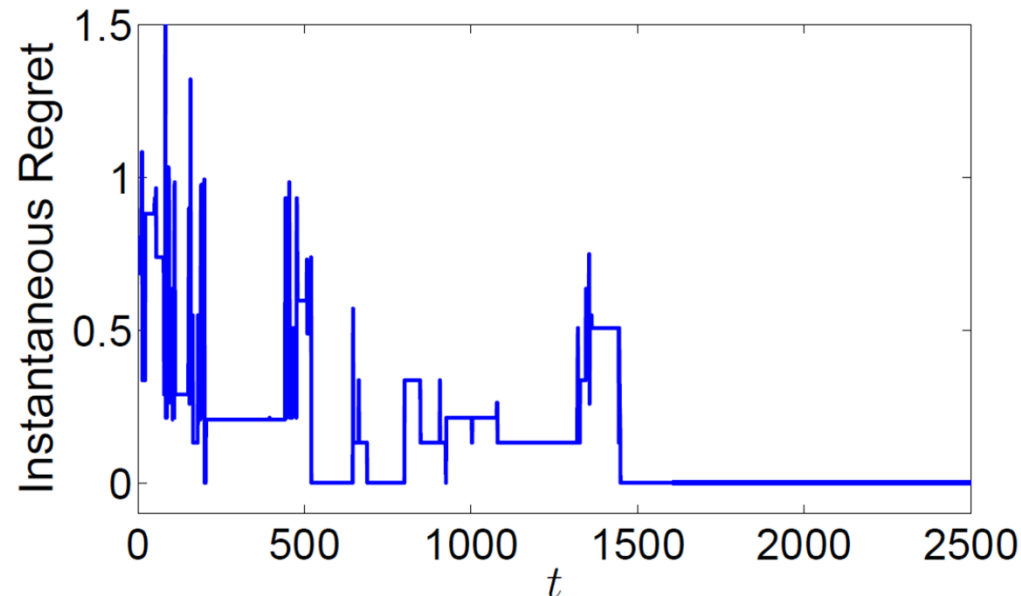
- 推广的置信上界法
 - 遗憾界: $O(d\sqrt{T})$

基于二元反馈的线性赌博机

- 支持更加常见的二元反馈
 - 是否购买、是否点击、是否喜欢
 - [Zhang et al., ICML 2016]

➤ 实验效果

■ $\mathcal{D} \subseteq \mathbb{R}^{10}$ and $|\mathcal{D}| = 100$

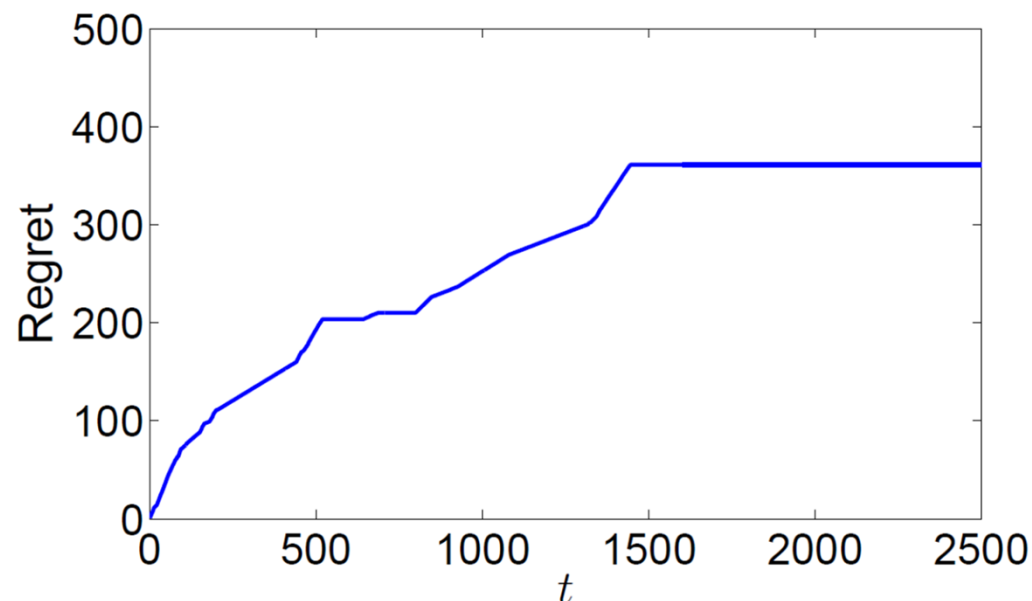


基于二元反馈的线性赌博机

- 支持更加常见的二元反馈
 - 是否购买、是否点击、是否喜欢
 - [Zhang et al., ICML 2016]

➤ 实验效果

■ $\mathcal{D} \subseteq \mathbb{R}^{10}$ and $|\mathcal{D}| = 100$



目录

➤ 研究背景

- 机器学习

- 大数据

➤ 在线学习

- 定义

- 完全信息在线学习

- 赌博机在线学习

➤ 总结与展望

总结

➤ 在线学习

- ❑ 在线处理数据，适用于大数据
- ❑ 有完备的理论保障（Regret）

➤ 完全信息在线学习

- ❑ 强反馈问题
- ❑ 在线分类、回归（线性、核）

➤ 赌博机在线学习

- ❑ 弱反馈问题
- ❑ 推荐系统（实数反馈、二元反馈）

➤ 在线非凸优化

- ❑ 当 $f_t(\cdot)$ 非凸时，缺乏理论保障
- ❑ 特殊的非凸函数 [Zhang et al., AAI 2015]

➤ 特征动态发生变化

- ❑ 医疗设备的更新、升级
- ❑ 寻找特征之间的关联 [Hou et al., NIPS 2017]

➤ 函数动态发生变化

- ❑ 动态遗憾

$$\sum_{t=1}^T f_t(\mathbf{w}_t) - \sum_{t=1}^T \min_{\mathbf{w} \in \mathcal{W}} f_t(\mathbf{w})$$

- ❑ 在线更新多次 [Zhang et al., NIPS 2017]

谢谢大家!

- **Zhang, L.**, Jin, R., Chen, C., Bu, J., and He, X. (2012).
Efficient online learning for large-scale sparse kernel logistic regression.
In Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI).
- **Zhang, L.**, Yi, J., Jin, R., Lin, M., and He, X. (2013).
Online kernel learning with a near optimal sparsity bound.
In Proceedings of the 30th International Conference on Machine Learning (ICML).
- **Zhang, L.**, Yang, T., Jin, R., and Zhou, Z.-H. (2015).
Online bandit learning for a special class of non-convex losses.
In Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI).
- **Zhang, L.**, Yang, T., Jin, R., Xiao, Y., and Zhou, Z.-H. (2016).
Online stochastic linear optimization under one-bit feedback.
In Proceedings of the 33rd International Conference on Machine Learning (ICML).

引用 (2)

- **Zhang, L.**, Yang, T., Yi, J., Jin, R., and Zhou, Z.-H. (2017).
Improved dynamic regret for non-degenerate functions.
In Advances in Neural Information Processing Systems 30 (NIPS).
- Hou, B.-J., **Zhang, L.**, and Zhou, Z.-H. (2017).
Learning with Feature Evolvable Streams.
In Advances in Neural Information Processing Systems 30 (NIPS).
- Rosenblatt, F. (1958).
The perceptron: a probabilistic model for information storage and organization in the brain.
Psychological Review, 65:386–407.
- Zinkevich, M. (2003).
Online convex programming and generalized infinitesimal gradient ascent.
In Proceedings of the 20th International Conference on Machine Learning (ICML).
- Shalev-Shwartz, S., Singer, Y., and Srebro, N. (2007).
Pegasos: primal estimated sub-gradient solver for SVM.
In Proceedings of the 24th International Conference on Machine Learning (ICML).

引用 (3)

- Shalev-Shwartz, S. (2011).
Online learning and online convex optimization.
Foundations and Trends in Machine Learning, 4(2):107–194.
- Auer, P., Cesa-Bianchi, N., and Fischer, P. (2002).
Finite-time analysis of the multiarmed bandit problem.
Machine Learning, 47(2-3):235–256.
- Dani, V., Hayes, T. P., and Kakade, S. M. (2008).
Stochastic linear optimization under bandit feedback.
In Proceedings of the 21st Annual Conference on Learning (COLT).