强化学习在Web安全领域的应用探索

兜哥 | 《Web安全之机器学习》作者

FIT · FREEBUF

个人简介

- 十年互联网企业安全建设经验
- 关注机器学习，威胁情报和企业安全建设
- 《Web安全之机器学习入门》《Web安全之深度学习实战》作者
- FreeBuf专栏作者
- 公众号《兜哥带你学安全》

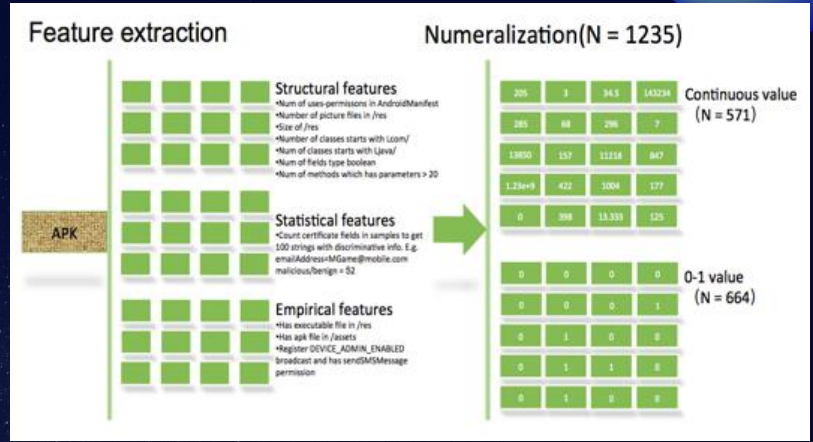
有监督学习在安全领域的应用

有监督学习在分类问题上，基于足量黑白样本，已经逐渐从实验室环境走向生产环境以及产品化

典型应用：

恶意软件检测、恶意APK检测、垃圾邮件检测、反欺诈

主要限制因素是如何获取足量的标记样本？

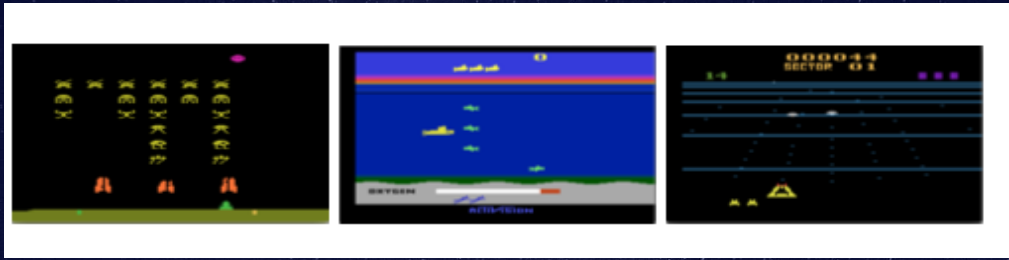


http://www.blackhat.com/eu-16/briefings.html#ai-based-antivirus-can-alphaav-win-the-battle-in-which-man-has-failed

强化学习受到众多关注

机器学习大致可以分为有监督学习，无监督学习与强化学习

强化学习可以解决多步决策问题



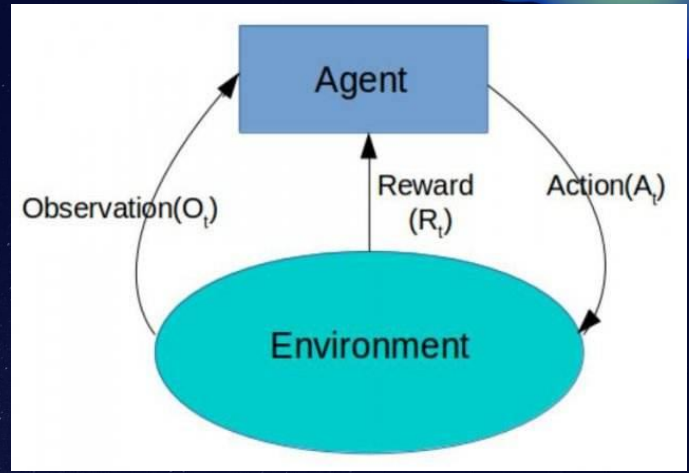
《 Playing Atari with Deep Reinforcement Learning 》，DeepMind

Match 4 - Google DeepMind Challenge Match: Lee Sedol vs AlphaGo

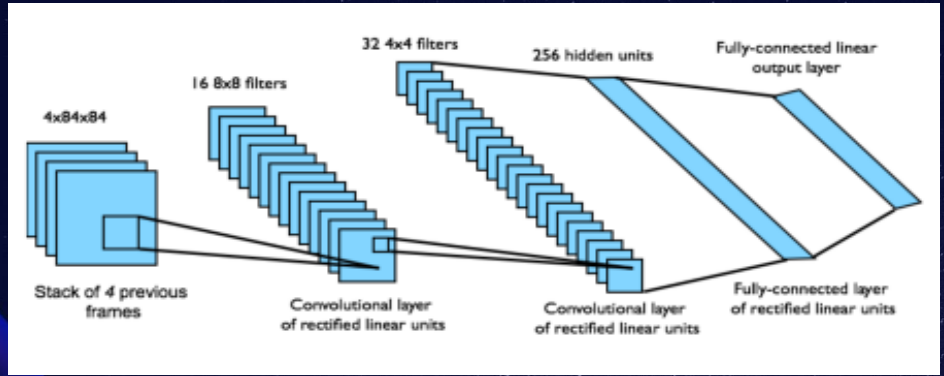
强化学习的概念

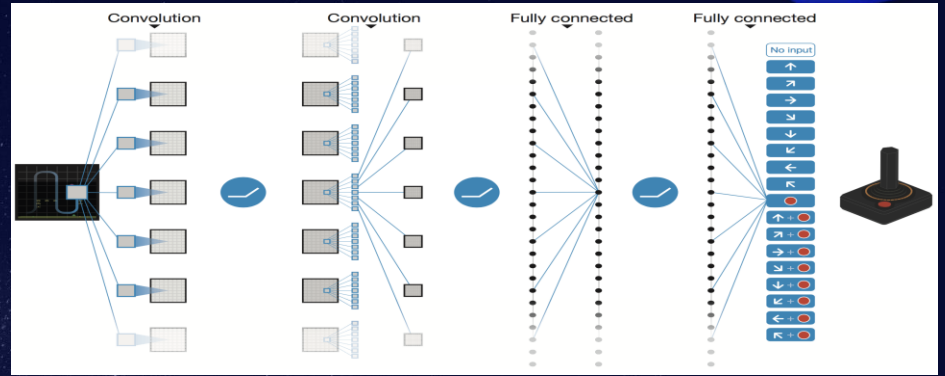
Environment，环境
Agent，智能体
Action，动作，全部动作成为动作空间
Observation，状态，包含Agent执行动作以后进入的下一个状态，全部状态成为状态空间
Reward，即所谓的奖励，Agent执行动作后会得到环境反馈的奖励
Q(Observation， Action)，可以理解为指定状态下执行指定动作的长期收益

智能体在与环境的交互中不断学习，在摸索中不断进步



DQN：强化学习+深度学习





《Human-level Control Through Deep Reinforcement Learning》,DeepMind

强化学习落地的关键要素

决定成败：

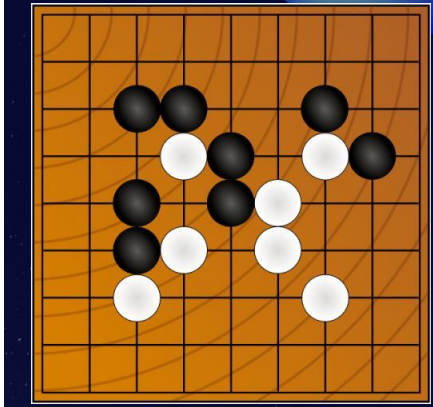
动作空间有限且明确

影响效果：

合适的深度神经网络结构（MLP、CNN还是ResNet）

策略选择算法（Q-Learning、Sarsa还是蒙特卡罗）

随机算法是贪婪算法还是e-贪婪算法



强化学习落地的关键要素

策略选择算法，或者理解为更新Q值的方式

其中α表示学习率，γ表示衰减因子。s_t和a_t分别表示当前的状态以及采取的动作，s_{t+1}表示下一个的状态，r_{t+1}表示当前状态采取动作后得到的回报。

Q-Learning算法：

$$Q_{k+1}(s_t,a_t) = Q_k(s_t,a_t) + \alpha(r_{t+1} + \gamma \max_a Q_k(s_{t+1},a) - Q_k(s_t,a_t))$$

Sarsa算法：

$$Q_{k+1}(s_t,a_t) = Q_k(s_t,a_t) + \alpha(r_{t+1} + \gamma Q_k(s_{t+1},a_{t+1}) - Q_k(s_t,a_t))$$

随机算法

贪婪算法：

永远选择Q值最大的动作

e-贪婪算法

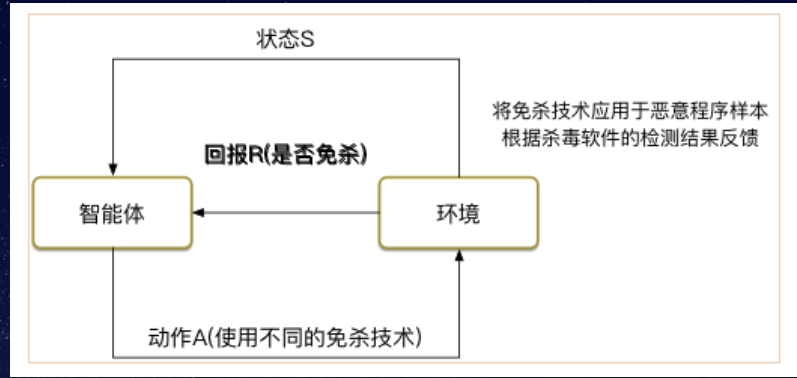
1-e的概率选择Q值最大的操作，但是以e的概率随机选择任意动作，具有冒险精神。e也可以动态调整，训练初期偏向冒险，后期偏向保守

强化学习的应用：恶意软件自动化免杀

Endgame公司在Blackhat2017提出了使用强化学习进行恶意软件自动化免杀，目前16%的样本可以达到自动化免杀

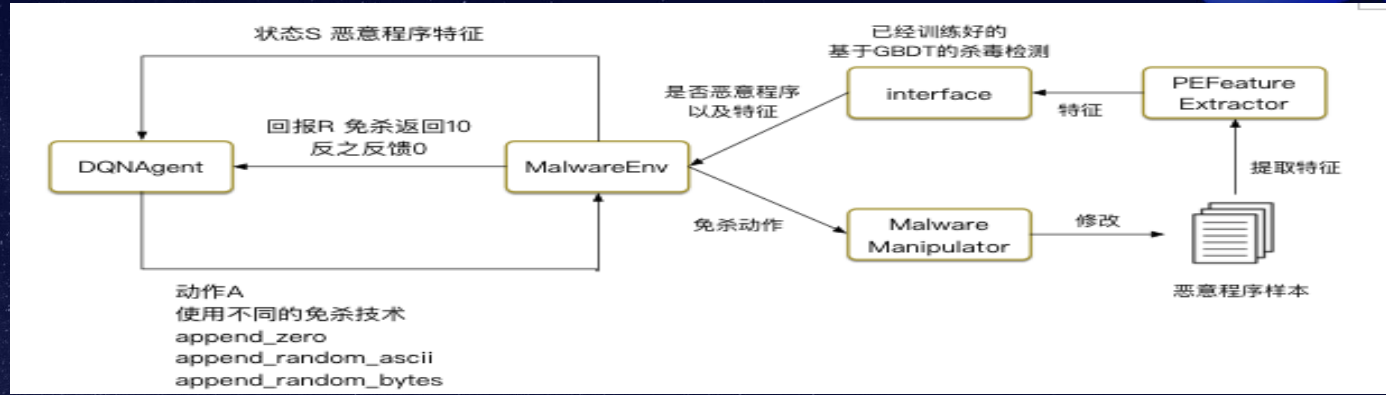
恶意软件的免杀操作可以归纳为简单几种，动作空间有限且明确，比如：

- 文件末尾追加随机内容
- 追加导入表
- 修改节名称
- 增加节



强化学习的应用：恶意软件自动化免杀

DQNAgent，智能体
MalwareEnv，强化学习环境
Interface，封装了杀毒软件的检测接口
MalwareManipulator，根据反馈对PE样本进行修改
PEFeature Extractor，从PE样本中提取特征



强化学习的应用：自动化测试WAF

以XSS为例，假设我们XSS样本为：
<IMG SRC=javascript:alert/1/>

常见的XSS绕过操作包括一下几种：
- 16进制编码

<IMG SRC=j&#x61vascript:alert/1/>
<IMG SRC=javascript:alert/1/>
<IMG SRC=javascript:alert/1/>
<IMG SRC=javascript:alert/1/>

- 10进制编码

<IMG SRC=j&#97vascript:alert/1/>
<IMG SRC=javascript:alert/1/>
<IMG SRC=javascript:alert/1/>
<IMG SRC=javascript:alert/1/>

- 插入注释

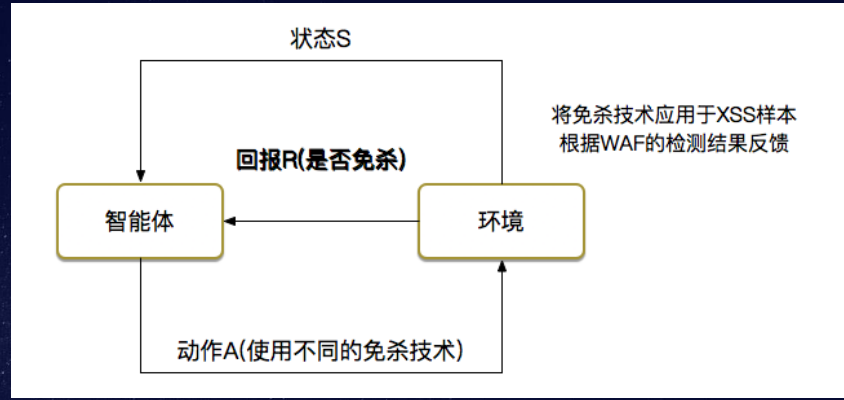
<IMG SRC=ja/*88888*/vascript:alert/1/>

强化学习的应用：自动化测试WAF

尝试使用强化学习进行WAF的自动化测试，以XSS绕过为例，目前的样本可以达到40%自动化免杀

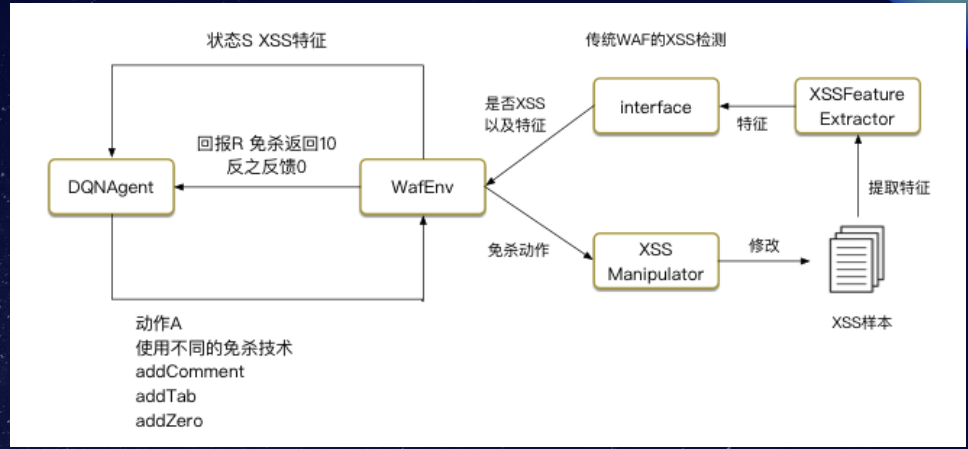
XSS的绕过操作可以归纳为简单几种，动作空间有限且明确，比如：

- 转换成16进制编码
- 转换成8进制编码
- 增加注释
- 增加Tab



强化学习的应用：自动化测试WAF

DQNAgent，智能体
WAFEnv，强化学习环境
Interface，封装了WAF的检测接口
XSSManipulator，根据反馈对XSS样本进行修改
XSSFeature Extractor，从XSS样本中提取特征



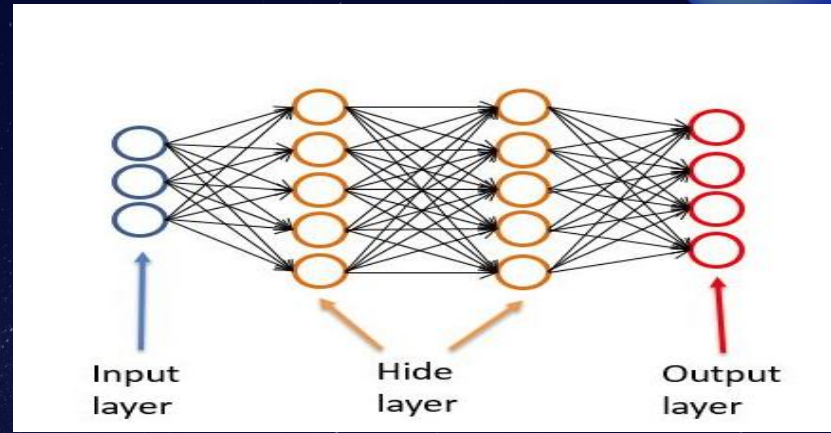
强化学习的应用：自动化测试垃圾邮件识别

数据集使用公开的安然数据集
特征提取使用词袋模型，检测模型使用MLP
准确率为96.24%，召回率为96.55%，漏报21个误报23个。

www2.aueb.gr/users/ion/data/enron-spam/

Contents of this directory:

- readme.txt
- Enron-Spam in pre-processed form:
  - Enron1
  - Enron2
  - Enron3
  - Enron4
  - Enron5
  - Enron6
- Enron-Spam in raw form:
  - ham messages:
    - beck-s
    - farmer-d
    - kaminski-v
    - kitchen-l
    - lokay-m
    - williams-w3
  - spam messages:
    - BG
    - GP
    - SH



强化学习的应用：自动化测试垃圾邮件识别

常见的绕过垃圾邮件检测的操作包括一下几种：

- **随机增加TAB**

thank you ,your ema
il address was obtained from a purchased list ,reference # 2020 mid = 3300 . if you wish to unsubscribe

- **大小写混淆**

thank you ,your email ADDRESS was obtained from a purchased list ,reference # 2020 mid = 3300 . if you wish to unsubscribe

- 随机增加连字符

thank you ,your email ad-d-ress was obtained from a purchased list ,reference # 2020 mid = 3300 . if you wish to unsubscribe

- **使用错别字**

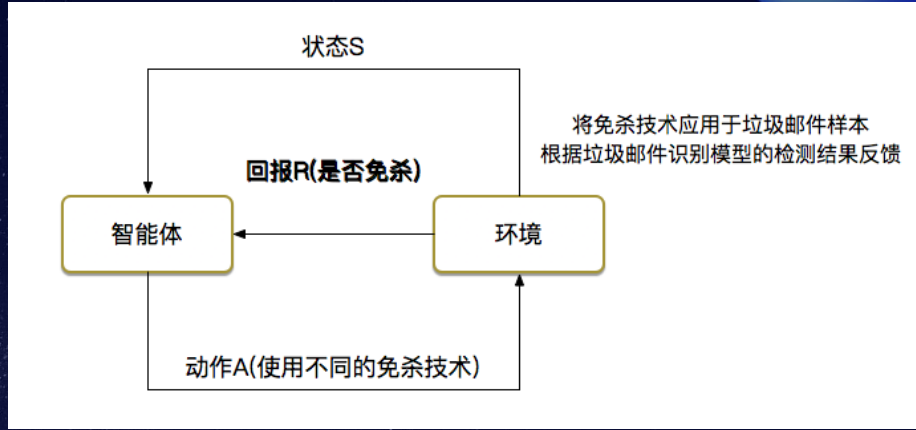
thank you ,your eemail addreess was obtained from a purchased list ,reference # 2020 mid = 3300 . if you wish to unsubscribe

强化学习的应用：自动化测试垃圾邮件识别

尝试使用强化学习自动化测试垃圾邮件识别，目前的样本可以达到16%自动化免杀

垃圾邮件的绕过操作可以归纳为简单几种，动作空间有限且明确，比如：

- 随机使用连字符
- 使用错别字
- 大小写混淆
- 增加Tab



强化学习的应用：自动化测试垃圾邮件识别

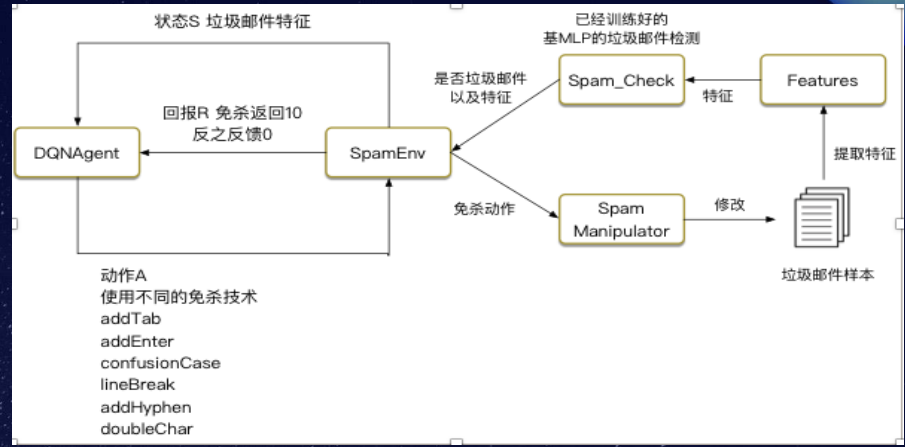
DQNAgent，智能体

SpamEnv，强化学习环境

Interface，封装了MLP垃圾邮件识别的检测接口

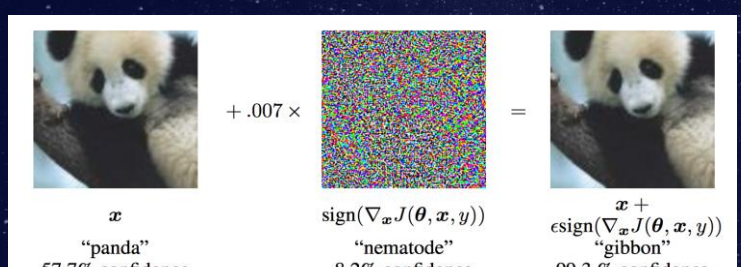
SpamManipulator，根据反馈对垃圾样本进行修改

Feature Extractor，从垃圾邮件样本中提取特征

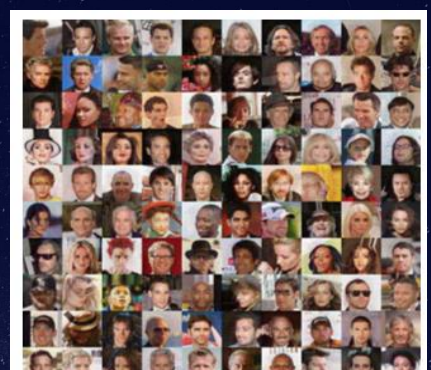


针对强化学习模型的攻击

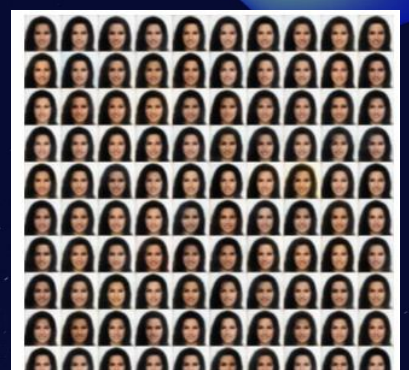
基于图像模型的攻击已经非常成熟，图像内容微小的变化就可以欺骗机器学习模型“指鹿为马“



把熊猫识别为长臂猿



对抗样本

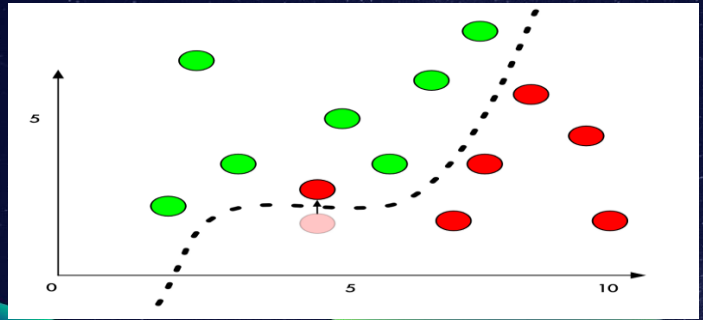


对抗样本还原成的图片

Jernej Kos, Ian Fischer, Dawn Song,《Adversarial examples for generative models 》

针对强化学习模型的攻击

攻击样本的本质就是针对特征向量在梯度方向适当的调整移动，最终跨越判断平面，导致误判。经典方法是Goodfellow提出的FGSM（Fast Gradient Sign Method）



Lets fool a binary linear classifier:

x	2	-1	3	-2	2	2	1	-4	5	1	← input example
w	-1	-1	1	-1	1	-1	1	1	-1	1	← weights
adversarial x	1.5	-1.5	3.5	-2.5	2.5	1.5	1.5	-3.5	4.5	1.5	

class 1 score before:
-2 + 1 + 3 + 2 + 2 - 2 + 1 - 4 - 5 + 1 = -3
=> probability of class 1 is 1/(1+e^(-(-3))) = 0.0474

$$P(y=1 \mid x;w,b) = \frac{1}{1+e^{-(w^T x+b)}} = \sigma(w^T x+b)$$

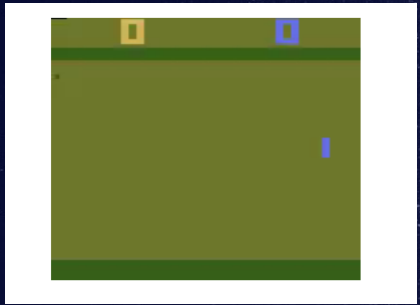
-1.5+1.5+3.5+2.5+2.5-1.5+1.5-3.5-4.5+1.5 = 2
=> probability of class 1 is now 1/(1+e^(-(2))) = 0.88
i.e. we improved the class 1 probability from 5% to 88%

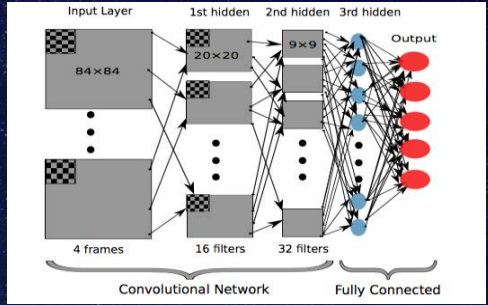
Fei-Fei Li & Andrej Karpathy & Justin Johnson　Lecture 9 - 72　3 Feb 2016

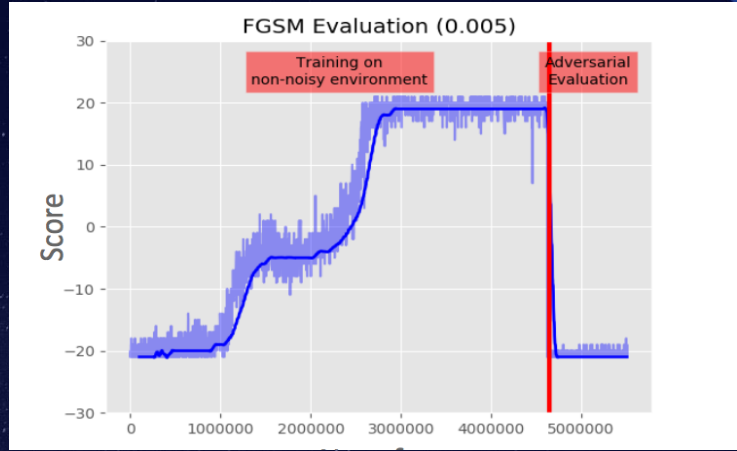
Fei-Fei Li，Stanford CS231n 2016

针对强化学习模型的攻击

攻击样本的思路同样可以用于攻击强化学习模型。本质上强化学习很多场景也会基于图像识别，针对图像的攻击方式同样生效。







Jernej Kos, Ian Fischer, Dawn Song,《Adversarial examples for generative models 》