线性分类器（1）

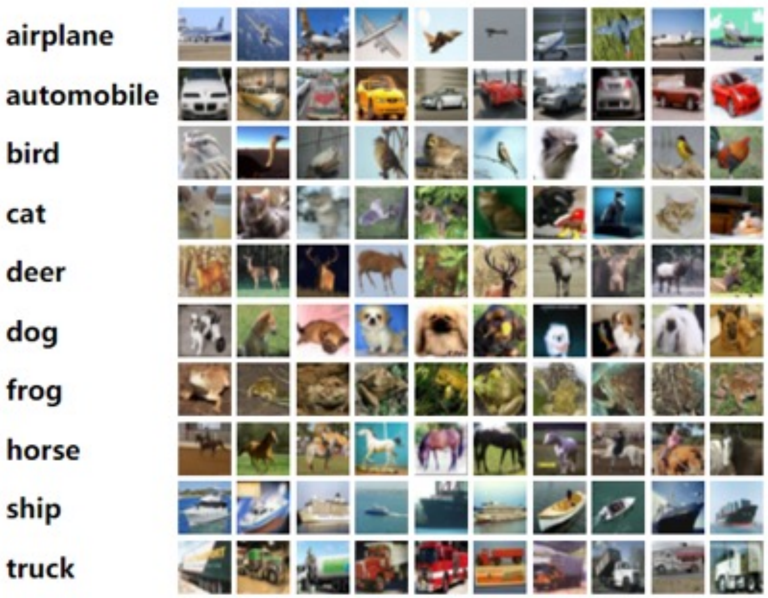
叶山 中国地质大学（北京）

yes@cugb.edu.cn

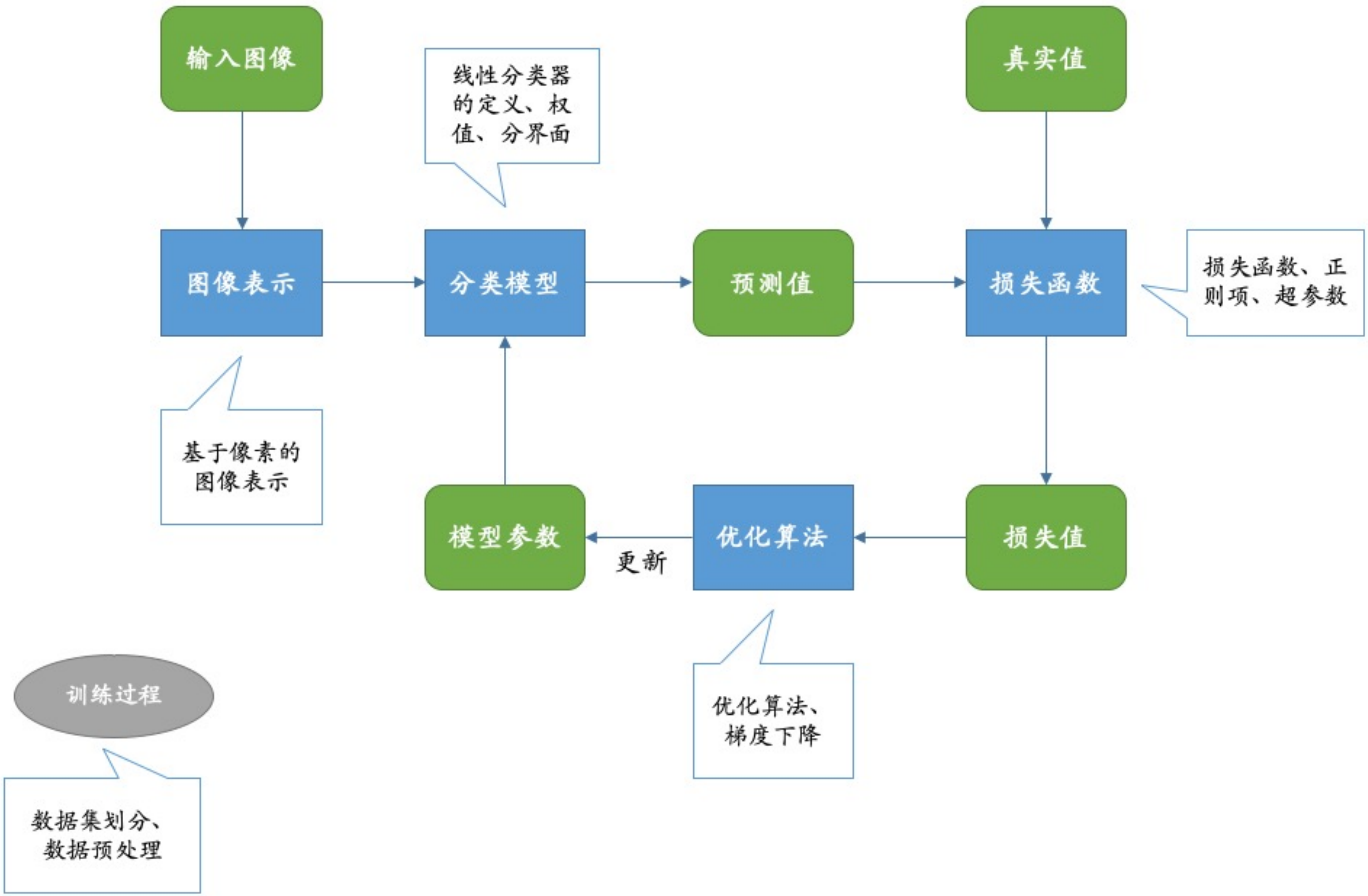
本教案部分参考以下课程：斯坦福大学 CS 231、密歇根大学 EECS 442 & 498、威斯康辛大学麦迪逊分校 CS 766

CIFAR-10 数据集

CIFAR = Canadian Institute For Advanced Research



CIFAR-10是一个常用的彩色图片分类数据集，包含了6万张32*32像素的彩色图像。这些图像被分为10个类别，包括飞机、汽车、鸟、猫、鹿、狗、青蛙、马、船和卡车。其中，有50000张图像被用于训练，剩下的10000张则用于测试。


输入图像
线性分类器的定义、权值、分界面
真实值
图像表示
分类模型
预测值
损失函数
损失函数、正则项、超参数
基于像素的图像表示
模型参数
更新
优化算法
损失值
训练过程
优化算法、梯度下降
数据集划分、数据预处理

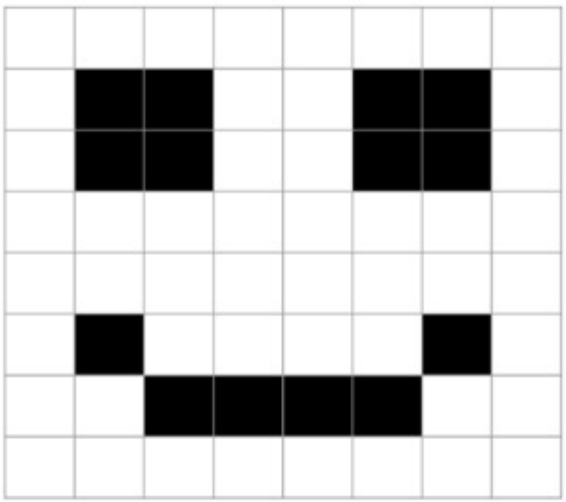
图像类型

二进制图 Binary

灰度图 Gray scale

RGB彩图

二进制图

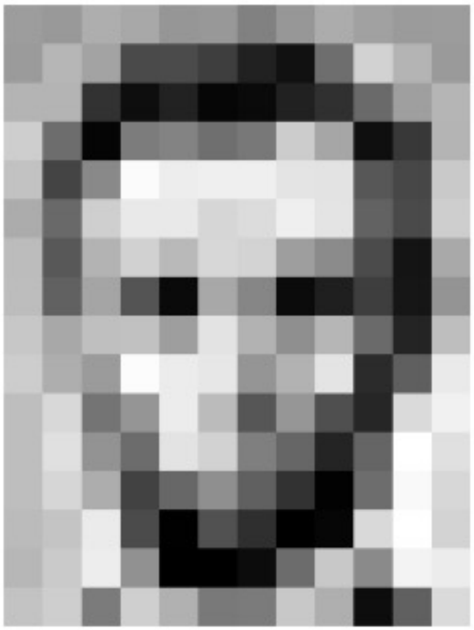


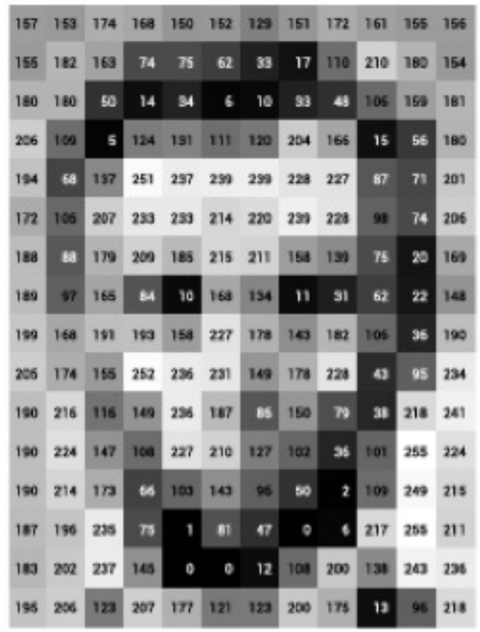
0	0	0	0	0	0	0	0
0	1	1	0	0	1	1	0
0	1	1	0	0	1	1	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	1	0	0	0	0	1	0
0	0	1	1	1	1	0	0
0	0	0	0	0	0	0	0

q

p

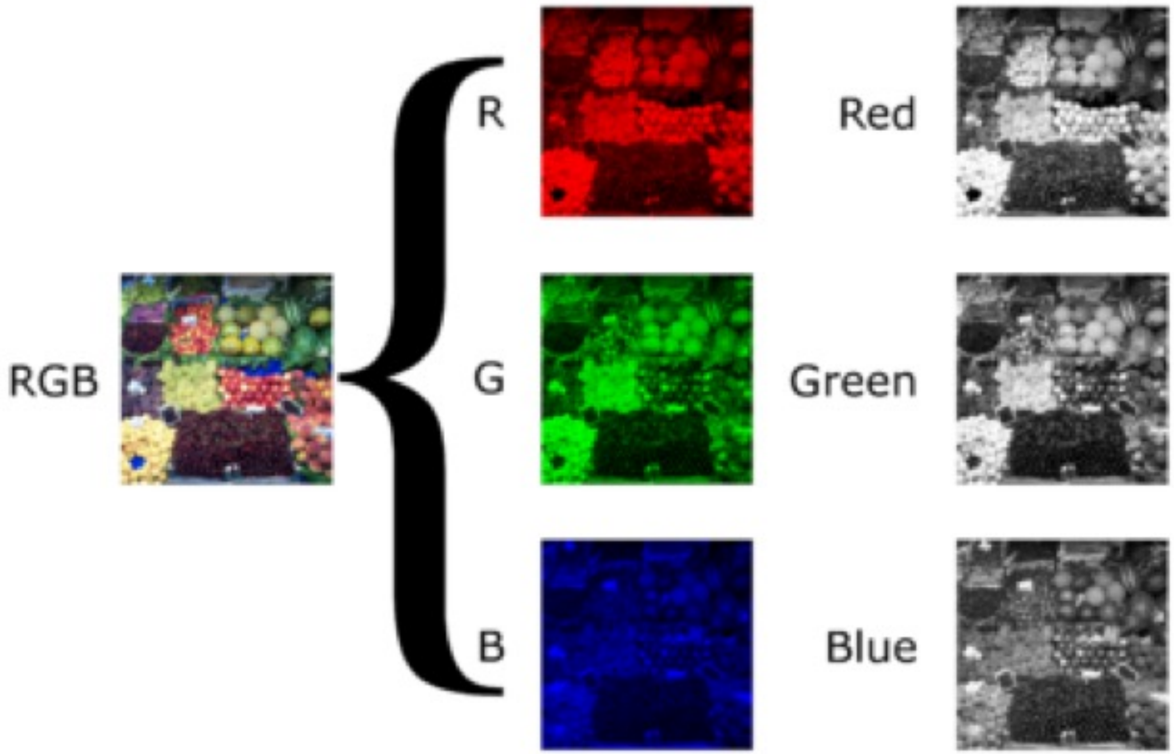
灰度图





157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

RGB彩图



图像表示

像素表示	全局特征表示	局部特征表示

线性分类器的输入是向量。

最简单的表示方法：将像素值依次展开为一个长向量（Flatten 操作）。

CIFAR-10图片展开后，向量的长度（维度）是多少？

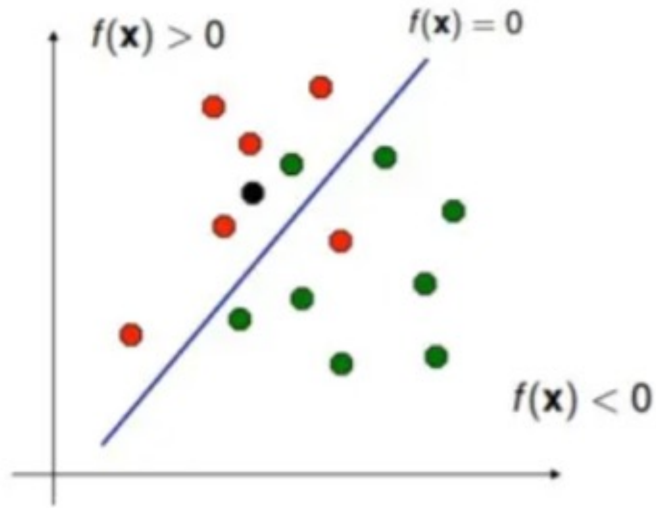
$32*32*3 = 3072$

分类器的选择

线性分类器

定义

- 线性分类器是一种基于线性方程的分类算法，它通过使用一个“超平面”来将不同类别的数据分隔开。在计算机视觉领域，它是一种将输入图像特征转化为类别分数的线性映射。
- 线性分类器使用一个或多个线性组合的特征来做出分类决策。在二维空间中，这个超平面可以是一条直线；在三维空间中，它是一个平面；而在更高维度的空间中，它是一个超平面。



特点

- 形式简单、容易理解。
- 是一些其他模型的基础：通过高维映射（支持向量机）或搭建层级结构（神经网络）可以形成功能强大的复杂模型。

线性分类器：代数视角

$$f_i(x, \omega_i) = \omega_i^T x + b_i$$

$$i = 1, 2, \ldots, c$$

- x代表输入的d维图像向量
- c为类别个数
- $\omega_i = [\omega_{i1}, \omega_{i2}, \ldots, \omega_{id}]^T$为第$i$个类别的权值向量，$b_i$为第$i$个类别的偏置
- 每个类都有自己的权值ω和偏置b

决策规则：

如果$f_i(x) > f_j(x), \forall i \neq j$

则把输入图像x归类到第i类

线性分类器：代数视角

任务：为图片分类，类别组包括狗、猫、鸟

$$f_i(x, \omega_i) = \omega_i^T x + b_i$$

$$i = 1, 2, 3$$

?

线性分类器：代数视角

图像表示x

56
231
24
2

（假设该图只有4个像素）

决策步骤：
1. 把图像展开为向量

$$f_i(x, \omega_i) = \omega_i^T x + b_i$$

$$i = 1, 2, 3$$

线性分类器：代数视角

$$
\begin{array}{c}
\text{权值 } \omega_i^T \\
\begin{array}{cc}
\text{狗 } \omega_1^T & \\
\text{猫 } \omega_2^T & \\
\text{鸟 } \omega_3^T &
\end{array}
\begin{bmatrix} 0.2 & -0.5 & 0.1 & 2 \\ 1.5 & 1.3 & 2.1 & 0 \\ 0 & 0.25 & 0.2 & -0.3 \end{bmatrix}
\end{array}
\cdot
\begin{array}{c}
\text{图像表示}x \\
\begin{bmatrix} 56 \\ 231 \\ 24 \\ 2 \end{bmatrix}
\end{array}
+
\begin{array}{c}
\text{偏置}b_i \\
\begin{array}{c} b_1 \\ b_2 \\ b_3 \end{array}
\begin{bmatrix} 1.1 \\ 3.2 \\ -1.2 \end{bmatrix}
\end{array}
=
\begin{array}{c}
\text{得分}f_i \\
\begin{array}{c} f_1 \\ f_2 \\ f_3 \end{array}
\begin{bmatrix} -98 \\ 435 \\ 63 \end{bmatrix}
\end{array}
$$

决策步骤:

1. 把图像展开为向量
2. 计算每个类别的分数

$$f_i(x, \omega_i) = \omega_i^T x + b_i$$

$$i = 1, 2, 3$$

线性分类器：代数视角

$$
\begin{array}{cc}
 & \text{权值 } \omega_i^T \\
\begin{array}{c} 狗\ \omega_1^T \\ 猫\ \omega_2^T \\ 鸟\ \omega_3^T \end{array} &
\begin{bmatrix} 0.2 & -0.5 & 0.1 & 2 \\ 1.5 & 1.3 & 2.1 & 0 \\ 0 & 0.25 & 0.2 & -0.3 \end{bmatrix}
\end{array}
\cdot
\begin{array}{c} \text{图像表示} x \\ \begin{bmatrix} 56 \\ 231 \\ 24 \\ 2 \end{bmatrix} \end{array}
+
\begin{array}{cc}
 & \text{偏置} b_i \\
\begin{array}{c} b_1 \\ b_2 \\ b_3 \end{array} & \begin{bmatrix} 1.1 \\ 3.2 \\ -1.2 \end{bmatrix}
\end{array}
=
\begin{array}{cc}
 & \text{得分} f_i \\
\begin{array}{c} f_1 \\ f_2 \\ f_3 \end{array} & \begin{bmatrix} -98 \\ 435 \\ 63 \end{bmatrix}
\end{array}
\checkmark
$$

这张图被划分为

猫

决策步骤：

1. 把图像展开为向量
2. 计算每个类别的分数
3. 查看得分最高的类别

$$f_i(x, \omega_i) = \omega_i^T x + b_i$$

$$i = 1, 2, 3$$

线性分类器：代数视角

权值 ω_i^T

狗 ω_1^T	0.2	-0.5	0.1	2	1.1
猫 ω_2^T	1.5	1.3	2.1	0	3.2
鸟 ω_3^T	0	0.25	0.2	-0.3	-1.2

·

图像表示x

56
231
24
2
1

=

得分f_i

f_1	-98
f_2	435
f_3	63

Bias Trick：将偏置吸收到权值里，在图像表示的末尾加一个等于1的维度。数学上等价。

这张图被划分为

猫

决策步骤：
1. 把图像展开为向量
2. 计算每个类别的分数
3. 查看得分最高的类别

$$f_i(x, \omega_i) = \omega_i^T x + b_i$$

$$i = 1, 2, 3$$

线性分类器：代数视角

权值 ω_i^T　　图像表示x　　偏置b_i　　得分f_i

$$\begin{array}{l} \text{狗}\ \omega_1^T \\ \text{猫}\ \omega_2^T \\ \text{鸟}\ \omega_3^T \end{array}\begin{bmatrix}0.2 & -0.5 & 0.1 & 2\\ 1.5 & 1.3 & 2.1 & 0\\ 0 & 0.25 & 0.25 & -0.3\end{bmatrix}\cdot\begin{bmatrix}56\\ 231\\ 24\\ 2\end{bmatrix}+\begin{array}{l}b_1\\ b_2\\ b_3\end{array}\begin{bmatrix}1.1\\ 3.2\\ -1.2\end{bmatrix}=\begin{array}{l}f_1\\ f_2\\ f_3\end{array}\begin{bmatrix}-98\\ 435\\ 63\end{bmatrix}$$

权值矩阵W　　偏置向量b　　得分向量f

线性分类器的矩阵表示：

$$f(x, W) = Wx + b$$

x代表输入图像，长度为像素数量d；
b为偏置向量，长度为类别个数c；
f为分数向量，长度为类别个数c；
W为权值矩阵。维度为$c * d$。

为CIFAR-10的图片做分类时，分类器的W、x、b的维度分别是多少？

x的长度（d）$= 3072 * 1$
b的长度（c）$= 10 * 1$
W的维度 $= 10 * 3072$

线性分类器：代数视角

$$\begin{array}{cc} & \text{权值}\ \omega_i^T \\ \begin{array}{c} \text{狗}\ \omega_1^T \\ \text{猫}\ \omega_2^T \\ \text{鸟}\ \omega_3^T \end{array} & \begin{bmatrix} 0.2 & -0.5 & 0.1 & 2 \\ 1.5 & 1.3 & 2.1 & 0 \\ 0 & 0.25 & 0.25 & -0.3 \end{bmatrix} \end{array} \cdot \begin{array}{c} \text{图像表示}x \\ \begin{bmatrix} 56 \\ 231 \\ 24 \\ 2 \end{bmatrix} \end{array} + \begin{array}{cc} & \text{偏置}b_i \\ \begin{array}{c} b_1 \\ b_2 \\ b_3 \end{array} & \begin{bmatrix} 1.1 \\ 3.2 \\ -1.2 \end{bmatrix} \end{array} = \begin{array}{cc} & \text{得分}f_i \\ \begin{array}{c} f_1 \\ f_2 \\ f_3 \end{array} & \begin{bmatrix} -98 \\ 435 \\ 63 \end{bmatrix} \end{array}$$

权值矩阵W　　偏置向量b　　得分向量f

忽略偏置 $f(x,W) = Wx$

$f(mx,W) = m \cdot Wx = mf(x,W)$

线性分类器的分类结果是线性的，可以线性缩放。

$$= \begin{array}{c} f_1 \\ f_2 \\ f_3 \end{array} \begin{bmatrix} -98 \\ 435 \\ 63 \end{bmatrix}$$

图像x

$$= \begin{array}{c} f_1 \\ f_2 \\ f_3 \end{array} \begin{bmatrix} -49 \\ 217.5 \\ 31.5 \end{bmatrix}$$

图像 $0.5x$

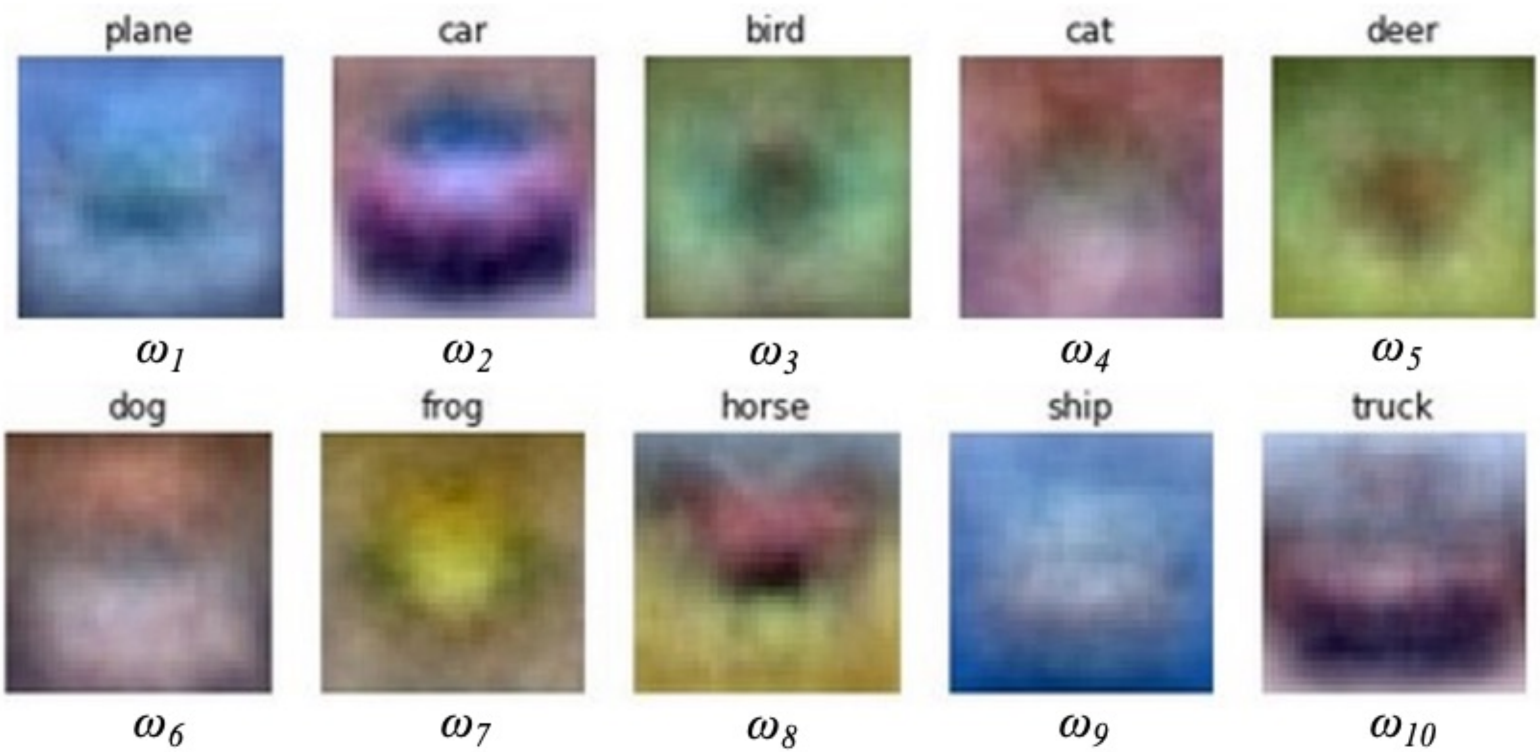
线性分类器：可视化视角

$f_i(x, \omega_i) = \omega_i^T x + b_i$

$i = 1, 2, \ldots, c$

$\omega_i = [\omega_{i1}, \omega_{i2}, \ldots, \omega_{id}]^T$为第$i$个类别的权值向量，$b_i$为第$i$个类别的偏置。

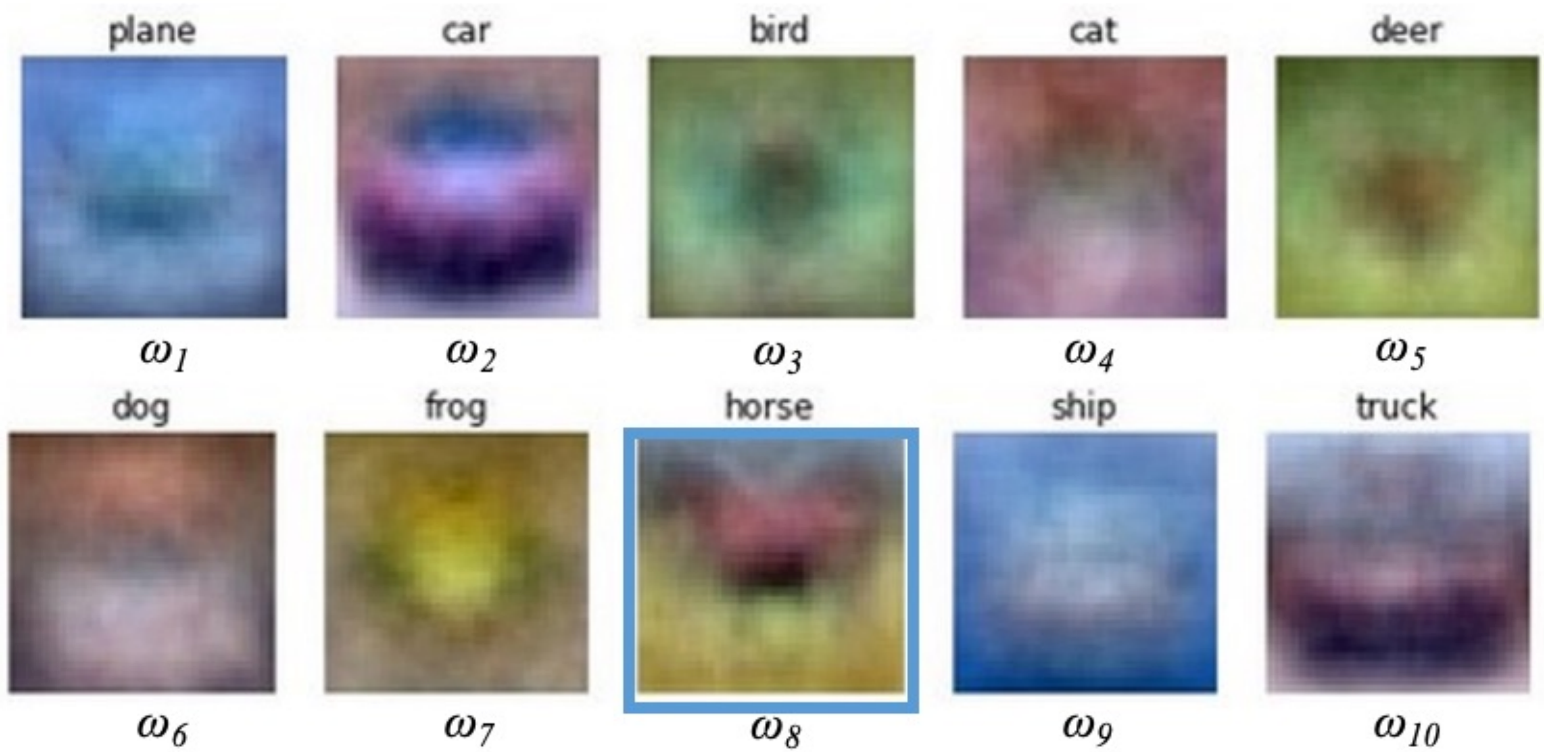
权值向量是记录该类图片宏观信息的模板。输入图像与模板的匹配度越高，点乘后得分越高。



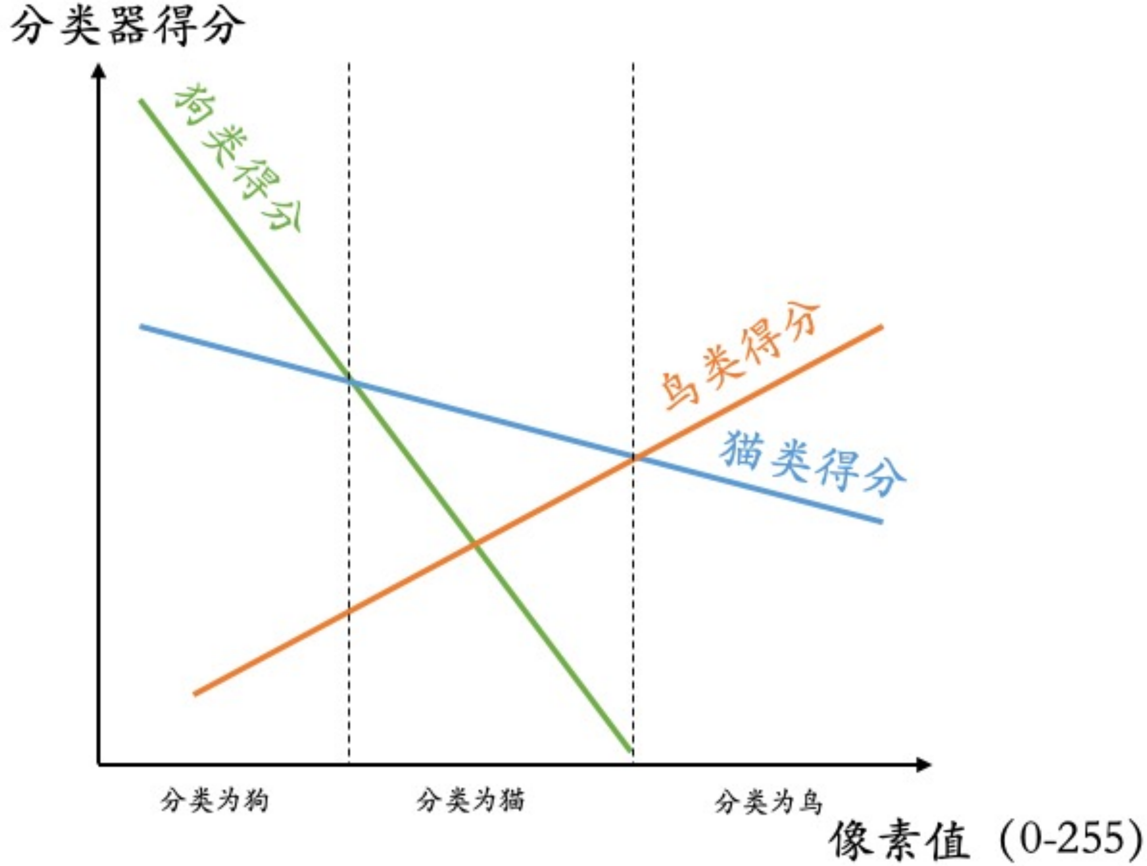
线性分类器：可视化视角

问题：单个模板无法捕获多种模式的数据。

- 例如：马类的权值模板有两个马头，因为数据集中，马类图片里的马头朝向各不相同。



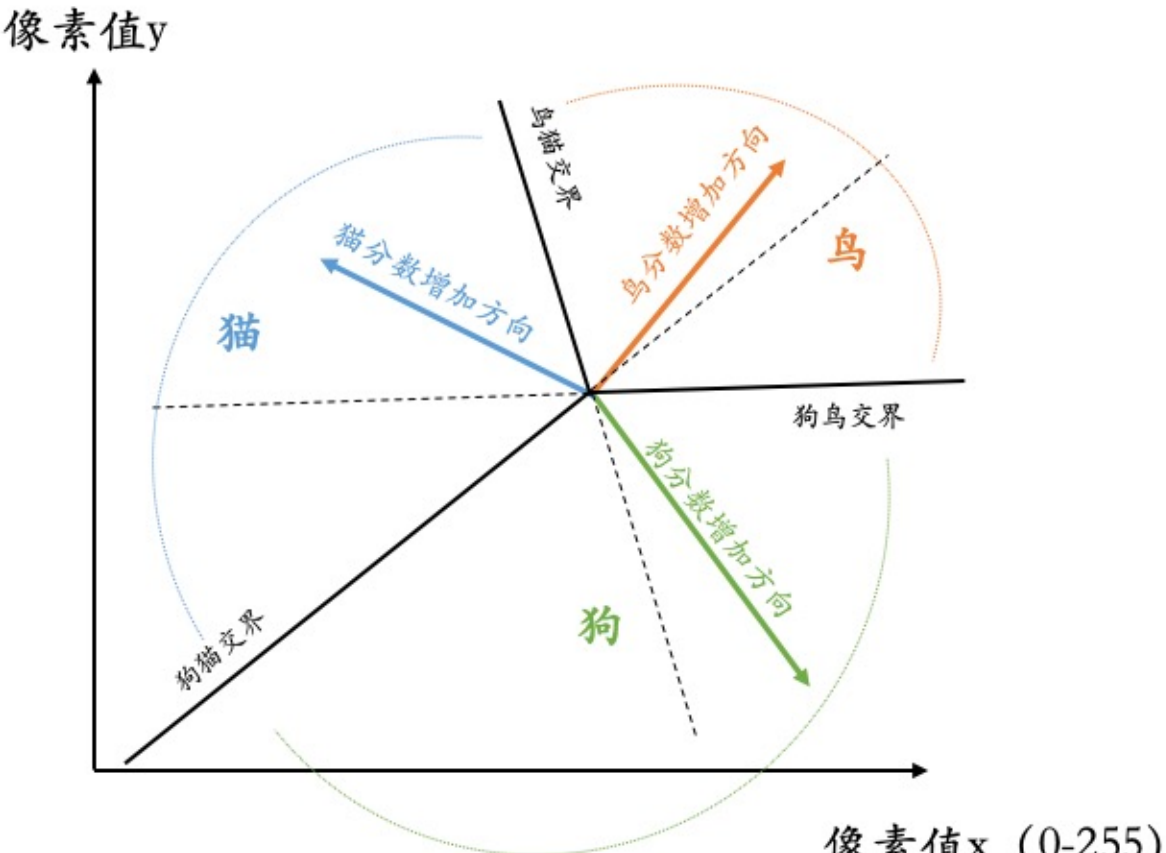
线性分类器：几何视角



只考虑1个像素的变化

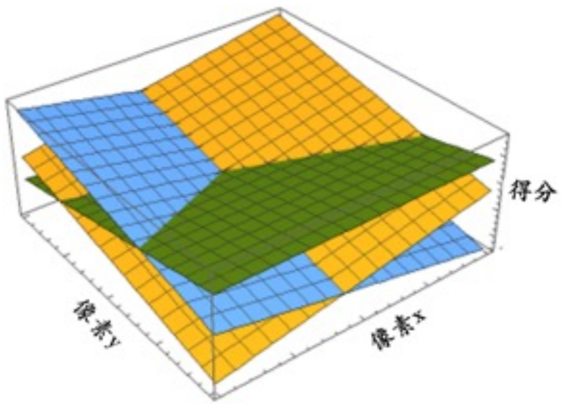
$$f(x,W) = Wx + b$$

线性分类器：几何视角

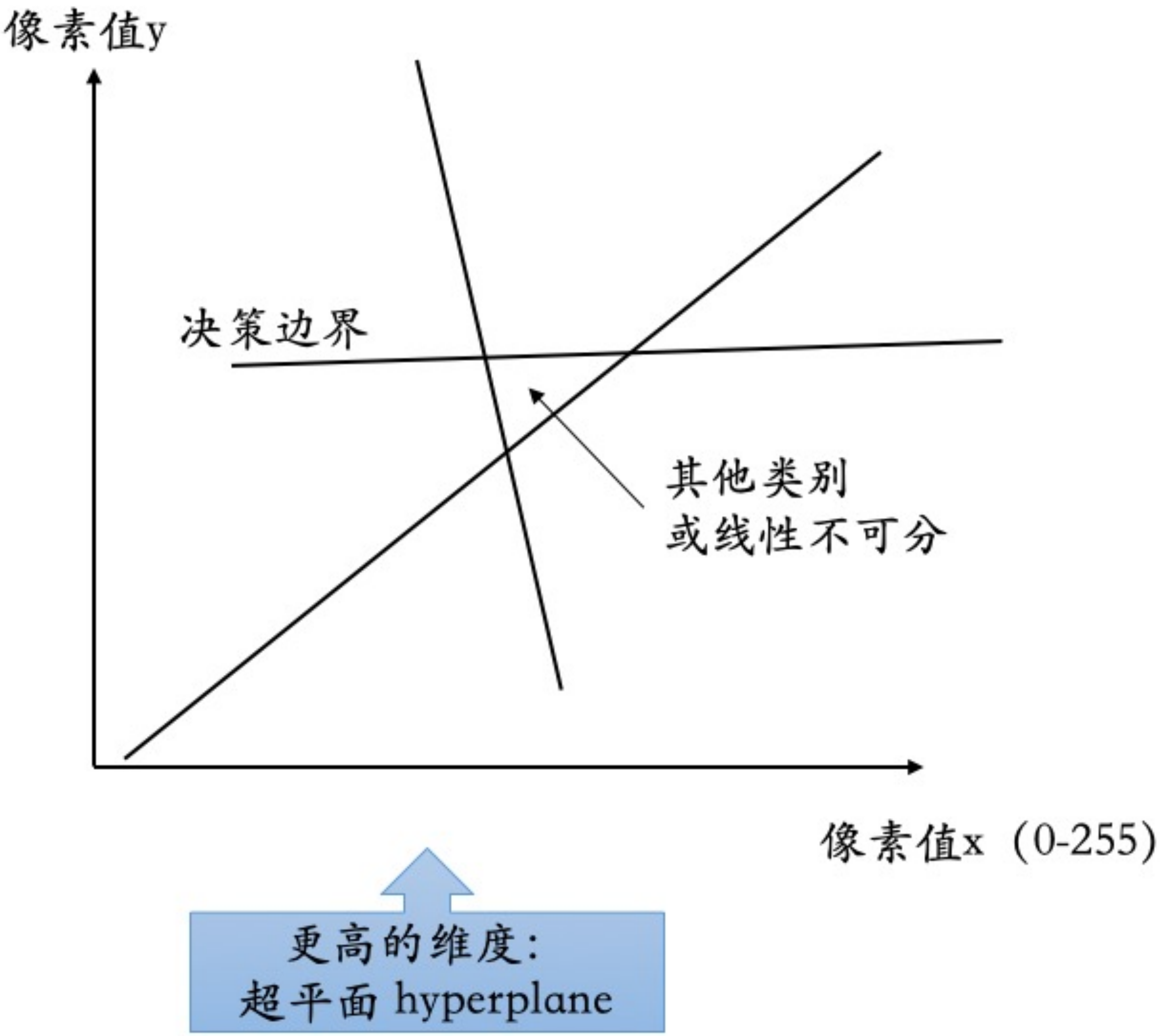


同时考虑2个像素的变化

$$f(x, W) = Wx + b$$

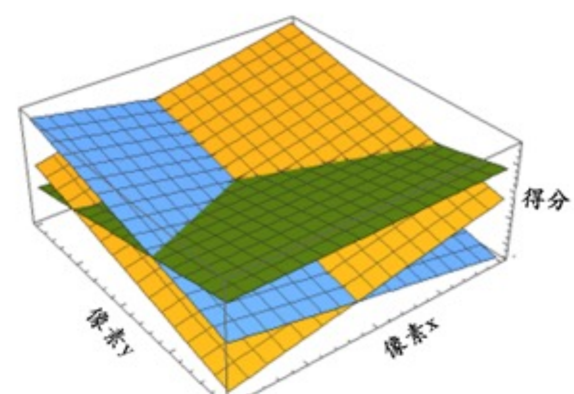


线性分类器：几何视角

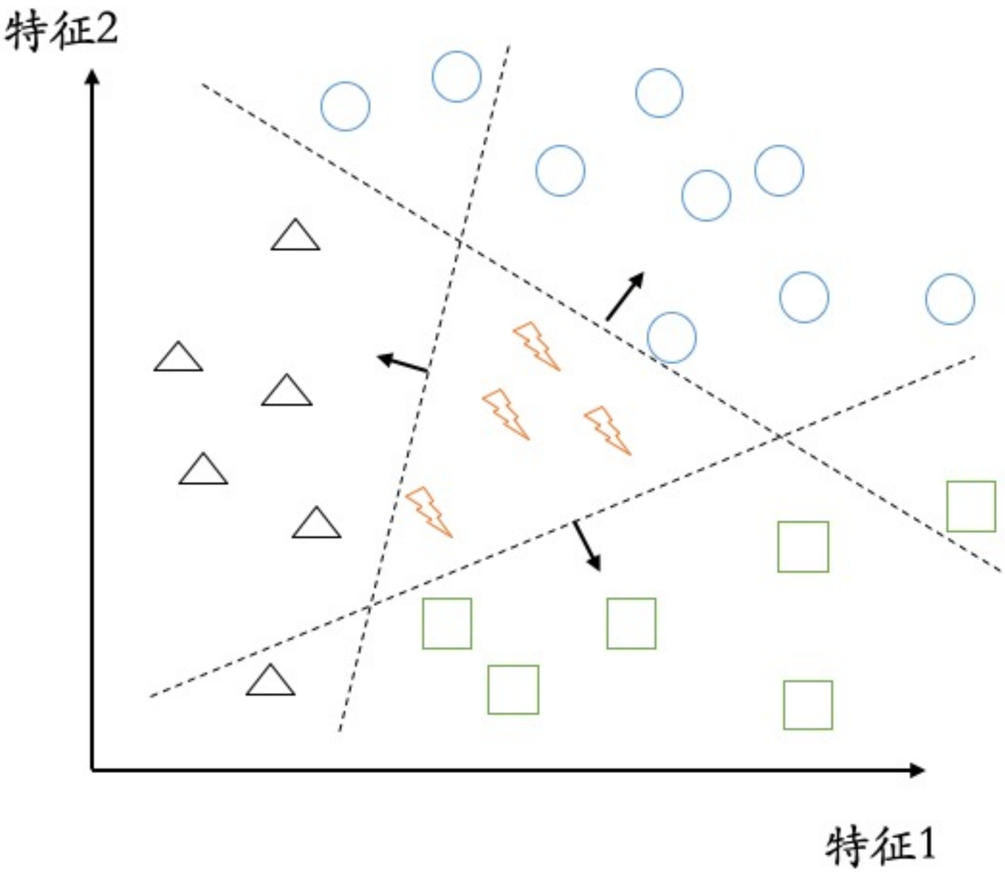


同时考虑2个像素的变化

$$f(x, W) = Wx + b$$



线性分类器：几何视角



分类器的核心任务:
寻找决策边界（超平面）
$\omega_i^T x + b_i = 0$

ω：控制正方向（类比斜率）
b：控制位移（类比截距）

箭头：分数增加最快的方向，即正方向，沿此方向，距离决策面越远，分数就越高。

小结

代数视角

如何线性地将图像映射为分值?

可视化视角

W权值矩阵是什么?

几何视角

什么是决策边界?

小结

airplane	-3.45	-0.51	3.42
automobile	-8.87	**6.04**	4.64
bird	0.09	5.31	2.65
cat	**2.9**	-4.22	5.1
deer	4.48	-4.19	2.64
dog	8.02	3.58	5.55
frog	3.78	4.49	**-4.34**
horse	1.06	-4.37	-1.5
ship	-0.36	-2.09	-4.79
truck	-0.72	-2.93	6.14

$$f(x, W) = Wx + b$$

定义一个权值矩阵W以及偏置向量b，可以通过线性分类器为一张图像x计算类别分数f。

存在的问题：初始化的W和b通常并非最佳，如何才能获得最佳的W和b？

答案：利用损失函数计算损失值，再用优化算法寻找让损失值最小的W和b。

损失函数

损失函数

- 损失函数的定义
- 多类支持向量机的损失
- 正则项和超参数。

损失函数

损失函数（loss function，又作代价函数cost function、客观性函数objective function）是衡量模型预测值与真实值之间差距的函数。

- 通过计算损失值，可以量化模型预测结果与真实结果之间的差距，从而评估模型的性能。
- 损失值越高，模型性能越差；损失值越低，模型性能越好。

损失函数

$$
\begin{array}{ll}
 & \text{权值 } \omega_i^T \\
\text{狗 } \omega_1^T & \begin{array}{|c|c|c|c|}\hline 0.2 & -0.5 & 0.1 & 2 \\ \hline 1.5 & 1.3 & 2.1 & 0 \\ \hline 0 & 0.25 & 0.2 & -0.3 \\ \hline \end{array} \\
\text{猫 } \omega_2^T & \\
\text{鸟 } \omega_3^T &
\end{array}
\cdot
\begin{array}{c}
\text{图像表示}x \\
\begin{array}{|c|}\hline 56 \\ \hline 231 \\ \hline 24 \\ \hline 2 \\ \hline \end{array}
\end{array}
+
\begin{array}{c}
\text{偏置}b_i \\
\begin{array}{c|c|}\cline{2-2} b_1 & 1.1 \\ \cline{2-2} b_2 & 3.2 \\ \cline{2-2} b_3 & -1.2 \\ \cline{2-2} \end{array}
\end{array}
=
\begin{array}{c}
\text{得分}f_i \\
\begin{array}{c|c|}\cline{2-2} f_1 & -98 \\ \cline{2-2} f_2 & 435 \\ \cline{2-2} f_3 & 63 \\ \cline{2-2} \end{array}
\end{array}
$$

分类器1： $f(x, W_1) = W_1 x + b_1$

$$
\begin{array}{ll}
 & \text{权值 } \omega_i^T \\
\text{狗 } \omega_1^T & \begin{array}{|c|c|c|c|}\hline 1 & 1.2 & 0.7 & 0 \\ \hline 0.3 & 0 & -0.4 & 1 \\ \hline 0.2 & 1.1 & 0.5 & 2 \\ \hline \end{array} \\
\text{猫 } \omega_2^T & \\
\text{鸟 } \omega_3^T &
\end{array}
\cdot
\begin{array}{c}
\text{图像表示}x \\
\begin{array}{|c|}\hline 56 \\ \hline 231 \\ \hline 24 \\ \hline 2 \\ \hline \end{array}
\end{array}
+
\begin{array}{c}
\text{偏置}b_i \\
\begin{array}{c|c|}\cline{2-2} b_1 & 0 \\ \cline{2-2} b_2 & -0.2 \\ \cline{2-2} b_3 & 0.2 \\ \cline{2-2} \end{array}
\end{array}
=
\begin{array}{c}
\text{得分}f_i \\
\begin{array}{c|c|}\cline{2-2} f_1 & 350 \\ \cline{2-2} f_2 & 9 \\ \cline{2-2} f_3 & 281 \\ \cline{2-2} \end{array}
\end{array}
$$

分类器2： $f(x, W_2) = W_2 x + b_2$

如何定量表示？

损失函数

- 损失函数反映了模型性能与模型参数之间的关联。它输出一个损失值（非负实数），该损失值就是对模型性能的描述。
- 损失值可以作为反馈信号来对分类器的参数进行调整，从而在下一轮里降低当前样本对应的损失值，提高分类器的分类效果。

损失函数

$$L=\frac{1}{N}\sum_{i=1}^{N}L_i(f(x_i,W),y_i)$$

L： 总损失值
L_i： 当前样本的损失值
$f(x_i,W)$： 当前样本的类别分数
y_i： 当前样本的真实标签（整数）
W： 权值（包含了偏置）
N： 样本的个数

多类支持向量机损失

$$s_{ij} = f_j(x_i, w_j, b_j) = w_j^T x_i + b_j$$

j指当前的类别编号，i指当前样本编号
w_j, b_j是第j个类别的分类器参数
s_{ij}是第i个样本在第j个类别中的预测分数

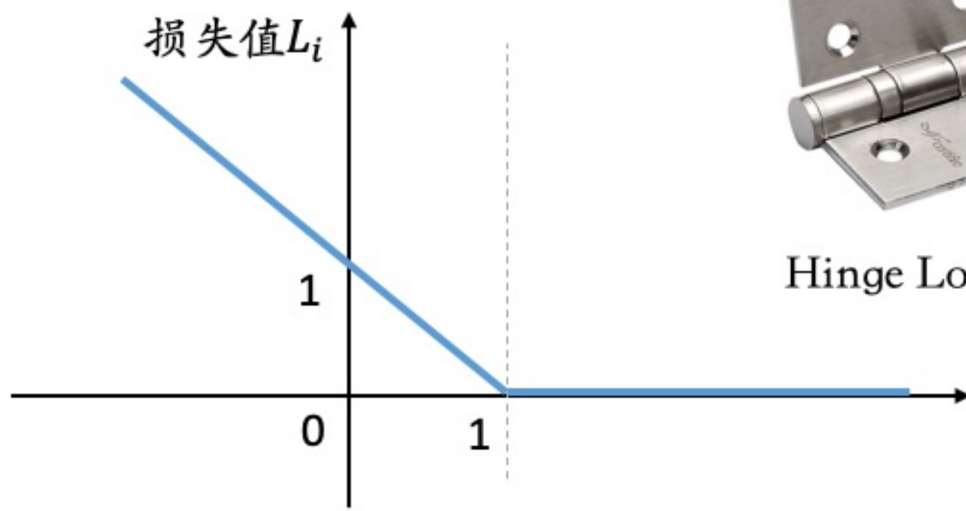
第i个样本的多类支持向量机损失:

$$L_i = \sum_{j \neq y_i} \begin{cases} 0 & \text{如果 } s_{y_i} \geq s_{ij} + 1 \\ s_{ij} - s_{y_i} + 1 & \text{其他情况} \end{cases}$$

$$= \sum_{j \neq y_i} \max(0, s_{ij} - s_{y_i} + 1)$$

s_{y_i}：第i个样本在真实类别里的预测分数

多类支持向量机损失



Hinge Loss 合页损失

第i个样本的多类支持向量机损失:

$$L_i = \sum_{j \neq y_i} \begin{cases} 0 & \text{如果 } s_{y_i} \geq s_{ij} + 1 \\ s_{ij} - s_{y_i} + 1 & \text{其他情况} \end{cases}$$

$$= \sum_{j \neq y_i} \max(0, s_{ij} - s_{y_i} + 1)$$

s_{y_i}: 第i个样本在真实类别里的预测分数

多类支持向量机损失

Fig1	Fig2	Fig3
		

	Fig1	Fig2	Fig3
cat	**3.2**	1.3	2.2
car	5.1	**4.9**	2.5
frog	-1.7	2.0	**-3.1**

加粗：s_{y_i}

$$L_i = \sum_{j \neq y_i} \max(0, s_{ij} - s_{y_i} + 1)$$

$$L_{Fig1} = \max(0, 5.1 - 3.1 + 1) + \max(0, -1.7 - 3.2 + 1) = 2.9$$

$$L_{Fig2} = \max(0, 1.3 - 4.9 + 1) + \max(0, 2.0 - 4.9 + 1) = 0$$

$$L_{Fig3} = \max(0, 2.2 - (-3.1) + 1) + \max(0, 2.5 - (-3.1) + 1) = 12.9$$