数字媒体技术基础

Meng Yang

www.smartllv.com

SUN YAT-SEN University

机器智能与先进计算教育部重点实验室

智能视觉语言学习研究组

目 录

4 媒体信息表示基础

Course Outline

目 录

4. 3变换空间表示方法

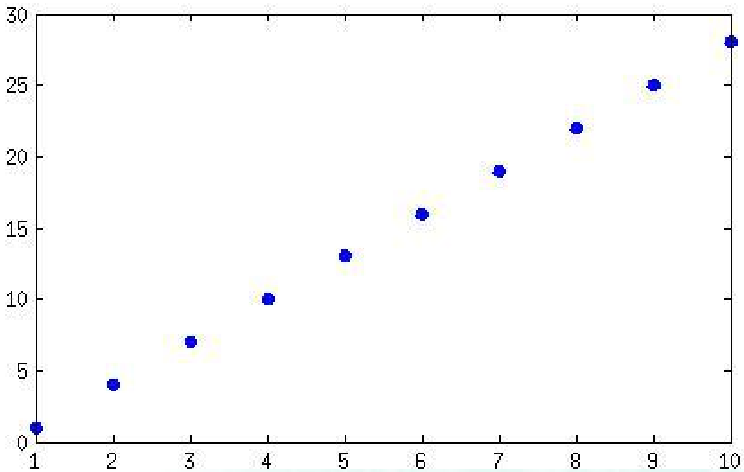
变换空间表示方法

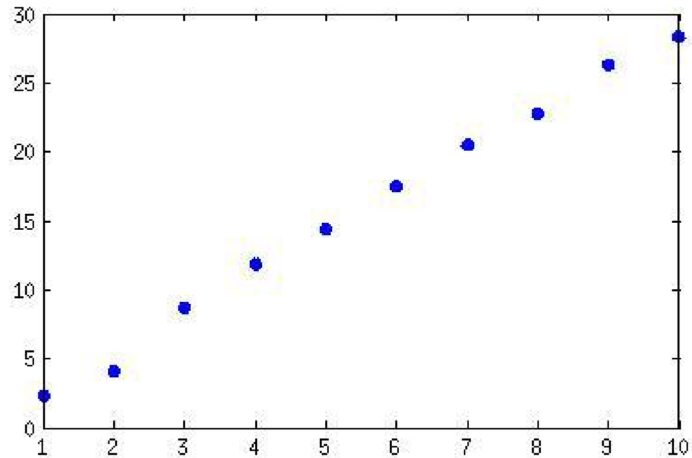
- 概述
  - 新表示 = 函数(原始表示)
- 主成分分析
- 频域分析
- 卷积分析

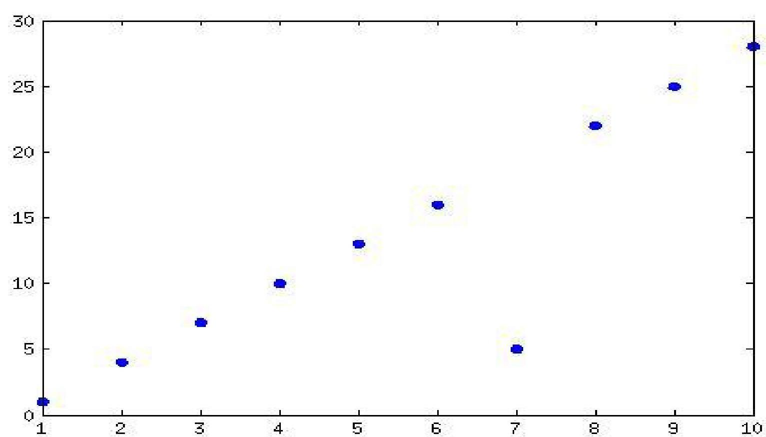
主成分分析

PCA基础

- 你的数据是多少维的？



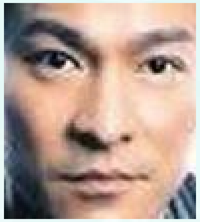




主成分分析

常见的数据特点

- 数据各维度之间不是互相独立的
  - 数据的内在维度(intrinsic dimensionality)通常远低于其表面维度
  - 因此，需要降低数据维度(dimensionality reduction)
  - PCA在降维方法中（可能）是最常用的一种



这是谁？

$96 \times 108 = 10368$?

主成分分析

- Starting point: 零阶表示
  - ✓ Zero-dimensional representation
  - ✓ 不允许使用任何维度，如何最佳表示$\boldsymbol{x}$?
  - ✓ 寻找某个固定(constant)的$\boldsymbol{m}$，使得

$$J_1(\boldsymbol{m}) = \min_{\boldsymbol{m}} \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{m}\|^2$$

  - ✓ 最优解：(证明？)

$$\boldsymbol{m}^* = \operatorname*{argmin}_{\boldsymbol{m}} \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{m}\|^2 = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$$

主成分分析

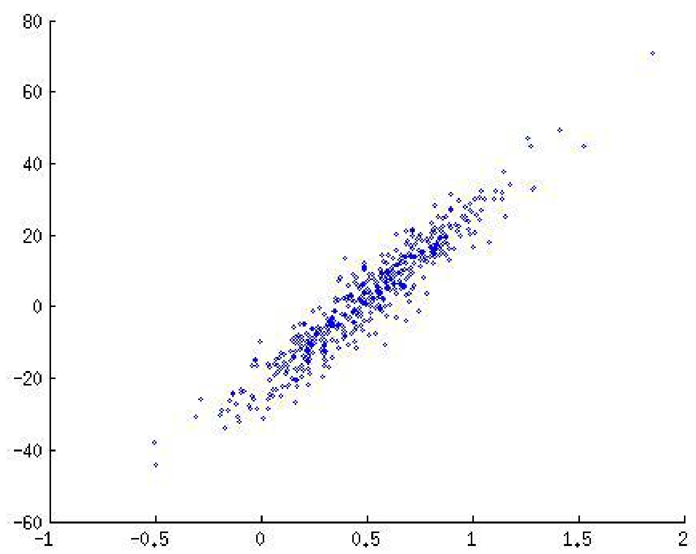
❑ 一维表示：数据维度间的线性关系

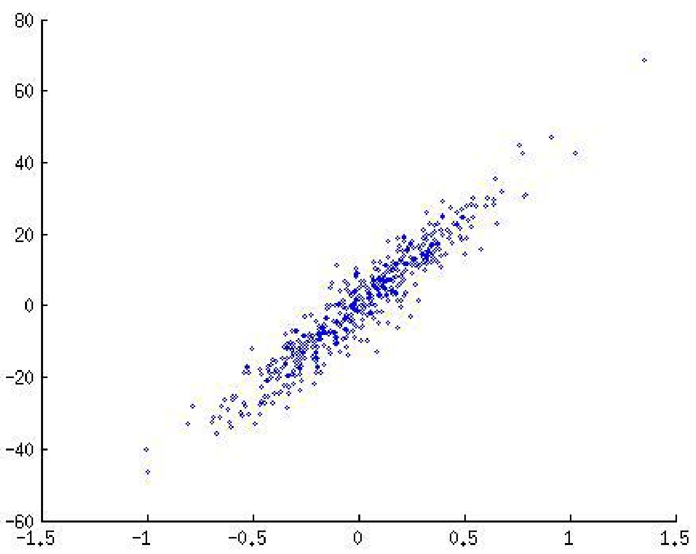
- 如同前面的例子
  - 数据是d维
  - 但是内在维度可能是m维的，$m < d$或者$m \ll d$
  - PCA用线性关系来降低维度

- $\boldsymbol{x} \in \mathbb{R}^d$：原来的高维数据（随机变量）
  - 训练样本：$\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$
- 假设$m = 1$，用原数据的单个线性组合来表示
  - $y_i = \boldsymbol{w}^T \boldsymbol{x}_i + b$
  - $y_1, y_2, \dots, y_n$ —— 新的数据/特征(features)
  - 如何寻找最佳的$\boldsymbol{w}$？如何找到最佳的b？

主成分分析

- Idea：选择什么方向？为什么？





✓ 什么方向最优？

主成分分析

形式化formalization：最大化方差

- ✓ 方差是衡量新特征包含信息多少的度量
  - 有时也称为能量energy
- ✓ 优化目标函数 $J_2(\boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{w}^T(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\right\|^2$
- ✓ 发现问题了吗?
  - $J_2(\boldsymbol{w})$可以是无穷大或者无穷小！
  - 最常用的解决办法：加上限制条件 $\|\boldsymbol{w}\|^2 = \boldsymbol{w}^T\boldsymbol{w} = 1$

$$\underset{\boldsymbol{w}}{\arg\max}\ \frac{1}{n}\sum_{i=1}^{n}\left\|(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T\boldsymbol{w}\right\|^2$$

$$\text{s.t.} \qquad \boldsymbol{w}^T\boldsymbol{w} = 1$$

  - s.t. –subject to,表示约束条件constraint(s)

主成分分析

- 简化simplification 变换transformation

$$\|(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T\boldsymbol{w}\|^2=\left((\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T\boldsymbol{w}\right)^T\left((\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T\boldsymbol{w}\right)$$
$$=\boldsymbol{w}^T(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T\boldsymbol{w}$$

$$\frac{1}{n}\sum_{i=1}^{n}\|(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T\boldsymbol{w}\|^2=\boldsymbol{w}^T\sum_{i=1}^{n}\frac{1}{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})^T\boldsymbol{w}$$
$$=\boldsymbol{w}^T\mathrm{Cov}(\boldsymbol{x})\boldsymbol{w}$$

主成分分析

优化optimization

- 拉格朗日乘子法 Lagrange multipliers
  - 将有约束的优化问题转化为无约束的优化问题
- Lagrangian 拉格朗日函数

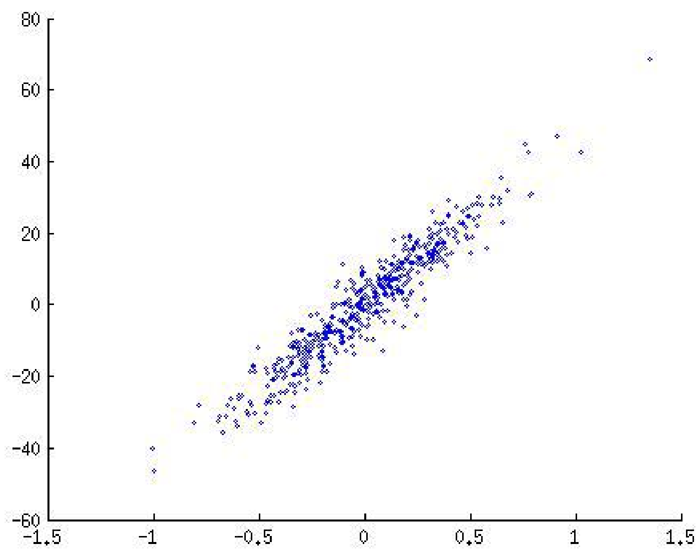
$$f(\boldsymbol{w},\lambda)=\boldsymbol{w}^T Cov(\boldsymbol{x})\boldsymbol{w}-\lambda(\boldsymbol{w}^T\boldsymbol{w}-1)$$

- λ：拉格朗日乘子Lagrange multiplier
- 最优的必要条件：$\frac{\partial f}{\partial \boldsymbol{w}}=\mathbf{0}$， $\frac{\partial f}{\partial \lambda}=0$
- $\frac{\partial f}{\partial \boldsymbol{w}}=2\mathrm{Cov}(\boldsymbol{x})\boldsymbol{w}-2\lambda\boldsymbol{w}=\mathbf{0}$
  - 我们这里的前提条件是什么？
  - 应该想到用哪一个公式？
- $\mathrm{Cov}(\boldsymbol{x})\boldsymbol{w}=\lambda\boldsymbol{w}$， $\boldsymbol{w}^T\boldsymbol{w}=1$！

主成分分析

- 选哪个特征向量？

  - ✓ $J_2(\boldsymbol{w}) = \boldsymbol{w}^T \mathrm{Cov}(\boldsymbol{x})\boldsymbol{w} = ?$
  - ✓ $\mathrm{Cov}(\boldsymbol{x})$是半正定的（如何证明？）
  - ✓ 选取λ_1（即最大特征值）对应的特征向量ξ_1为$\boldsymbol{w}_1$
    - 为什么？约束条件满足了吗？
  - ✓ 怎样用$\boldsymbol{w}_1$来近似$\boldsymbol{x}$？
    - 投影！
    - $\boldsymbol{x} \approx \overline{\boldsymbol{x}} + (\boldsymbol{w}_1^T(\boldsymbol{x} - \overline{\boldsymbol{x}}))\boldsymbol{w}_1$
    - 所以，$y_i = \boldsymbol{w}_1^T(\boldsymbol{x} - \overline{\boldsymbol{x}})$
    - 那么，$b = ?$



主成分分析

J_1和J_2的等价关系

- 若干向量
  - $\boldsymbol{x}_i$：降维之前的向量
  - $\boldsymbol{w}_1^T(\boldsymbol{x}_i - \overline{\boldsymbol{x}})\boldsymbol{w}_1 = y_i\boldsymbol{w}_1$：降维之后的向量
  - $\widehat{\boldsymbol{x}}$：在原空间中重建的向量
  - 目前的重建关系：$\widehat{\boldsymbol{x}}_i \approx \overline{\boldsymbol{x}} + y_i\boldsymbol{w}_1$
- J_1的目的是使得$\widehat{\boldsymbol{x}}_i$和$\boldsymbol{x}_i$尽可能相差小($\overline{\boldsymbol{x}}$固定为均值)
  - $J_1(\boldsymbol{w}, \boldsymbol{a}) = \sum_{i=1}^{n} \frac{1}{n} \|\boldsymbol{x}_i - (\overline{\boldsymbol{x}} + a_i\boldsymbol{w})\|^2$
  - $\boldsymbol{w}$：投影方向，a_i：投影系数
- 最小化J_1得到的a_i和$\boldsymbol{w}$与J_2得到的结果完全一致！
  - 试着去证明!

主成分分析

如果需要更多投影方向?

- What if we need $\boldsymbol{w}_2, \boldsymbol{w}_3, \dots$
  - 新的投影方向需要继续保持“能量”
  - 但是需要限制
  - $\boldsymbol{w}_2 \perp \boldsymbol{w}_1$， $\boldsymbol{w}_3 \perp \boldsymbol{w}_2$， $\boldsymbol{w}_3 \perp \boldsymbol{w}_1$， $\dots$
- 在上述限制条件下
  - $\boldsymbol{w}_2 = \boldsymbol{\xi}_2$， $\boldsymbol{w}_3 = \boldsymbol{\xi}_3$， $\dots$
  - 重建系数: $\boldsymbol{w}_j^T(\boldsymbol{x} - \overline{\boldsymbol{x}})$
- 总之，

$$\boldsymbol{x} \approx \overline{\boldsymbol{x}} + (\boldsymbol{w}_1^T(\boldsymbol{x} - \overline{\boldsymbol{x}}))\boldsymbol{w}_1 + (\boldsymbol{w}_2^T(\boldsymbol{x} - \overline{\boldsymbol{x}}))\boldsymbol{w}_2 + \cdots$$

主成分分析

重建和原数据的关系

- 假设$n > d$，即数据比维数多
  - 进一步假设$\mathrm{Cov}(\boldsymbol{x})$可逆
  - 如果$n < d$，那么情况如何、还能做PCA变换吗？
- $\mathrm{Cov}(\boldsymbol{x})$是$d \times d$矩阵，有$d$个互相垂直的特征向量$\boldsymbol{\xi}_i$
  - 重建会有d个互相垂直的投影方向$\boldsymbol{w}_i$

$$\forall \boldsymbol{x}, \qquad \boldsymbol{x} = \bar{\boldsymbol{x}} + \sum_{i=1}^{d} \left(\boldsymbol{w}_i^T(\boldsymbol{x} - \bar{\boldsymbol{x}})\right)\boldsymbol{w}_i$$

  - 将$\boldsymbol{w}_i$拼成矩阵形式$W = [\boldsymbol{w}_1 \ \ \boldsymbol{w}_2 \ \ \cdots \boldsymbol{w}_d]$（$d \times d$）
  - 投影系数是$W^T(\boldsymbol{x} - \bar{\boldsymbol{x}})$，投影方向是$W$
  - $\boldsymbol{x} = \bar{\boldsymbol{x}} + WW^T(\boldsymbol{x} - \bar{\boldsymbol{x}})$ （为什么？）
  - 重建是完全精确的（没有误差），为什么？

主成分分析

降维

- 很多时候，有些投影方向是噪声
  - 需要扔掉一些方向
  - 扔掉哪些？扔掉多少？
- 去掉特征值最小的那些
- 通常保持90%的能量

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_T}{\lambda_1 + \lambda_2 + \cdots + \lambda_d} > 0.9$$

  - 寻找第一个T，使得上面的不等式成立

主成分分析

❑ 降维的损失

- ✓ 现在$\widehat{W} = [\boldsymbol{w}_1 \ \ \boldsymbol{w}_2 \ \ \cdots \boldsymbol{w}_T]$（$d \times T$）
- ✓ $\boldsymbol{x} - \widehat{\boldsymbol{x}} = \sum_{j=T+1}^{d} (\boldsymbol{w}_j^T(\boldsymbol{x} - \overline{\boldsymbol{x}}))\boldsymbol{w}_j = \sum_{j=T+1}^{d} \boldsymbol{e}_j$
  - 这个误差多大？
  - $\boldsymbol{e}_j^T \boldsymbol{e}_k = 0$ 如果 $j \neq k$（为什么？）
  - $\mathrm{E}(\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|^2) = \sum_{j=T+1}^{d} \mathrm{E}(\|\boldsymbol{e}_j\|^2)$（为什么？）
  - $\mathrm{E}(\|\boldsymbol{e}_j\|^2) = \lambda_j$ （为什么？）
- ✓ 这样降维保证平均（期望）重建误差最小
  - 直接优化重建误差J_1得到同样的结果

主成分分析

小结：PCA变换的步骤

- 训练样本：$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$
- 计算得到$\overline{\boldsymbol{x}}$和$\mathrm{Cov}(\boldsymbol{x})$
- 求得$\mathrm{Cov}(\boldsymbol{x})$的特征值和特征向量
  - Matlab，R，octave，…
- 根据特征值选定T
- 根据T的值确定矩阵$\widehat{W}$

- 对任何数据$\boldsymbol{x}$，其新的经过PCA变换得到的特征是

$$\boldsymbol{y} = \widehat{W}^T(\boldsymbol{x} - \overline{\boldsymbol{x}})$$

重建则为 $\boldsymbol{x} \approx \widehat{\boldsymbol{x}} = \overline{\boldsymbol{x}} + \widehat{W}\boldsymbol{y}$

频域分析

- 对于二维信号，其频谱的表示如下，其中
  - 高频部分代表细节、边缘和噪声
  - 低频占据绝大多少能 量，其中直流分量(零频)能量占比最 大。
  - 频率分布具有中心对 称性
- 傅里叶变换公式

Fourier transform (FT) (***Fourier integral***)

$$\boxed{X(\mathrm{j}\omega)=\int_{-\infty}^{\infty}x(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t}$$

Analysis equation

To obtain the spectrum

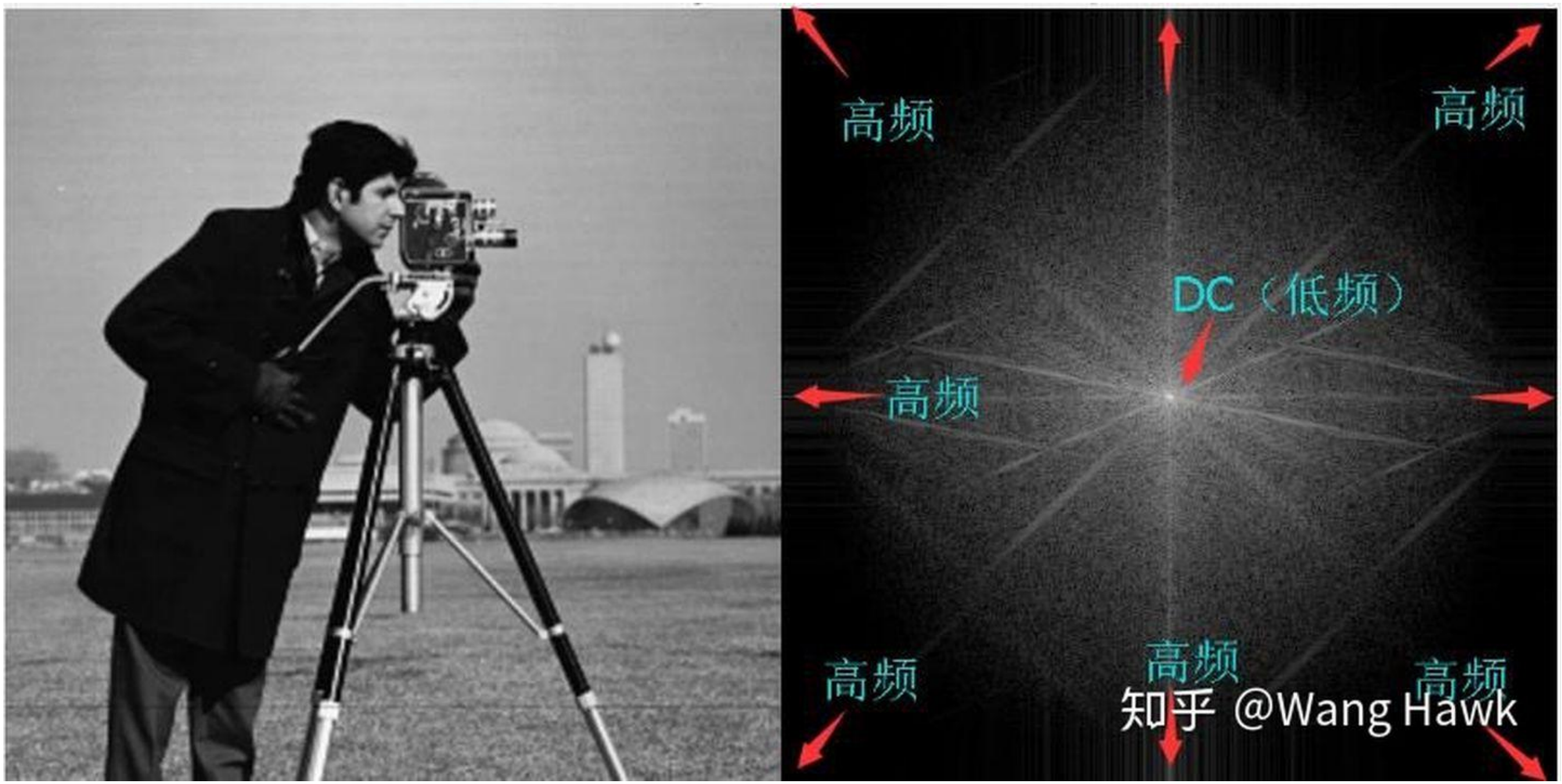
Inverse Fourier transform

$$\boxed{x(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}X(\mathrm{j}\omega)\mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{d}\omega}$$

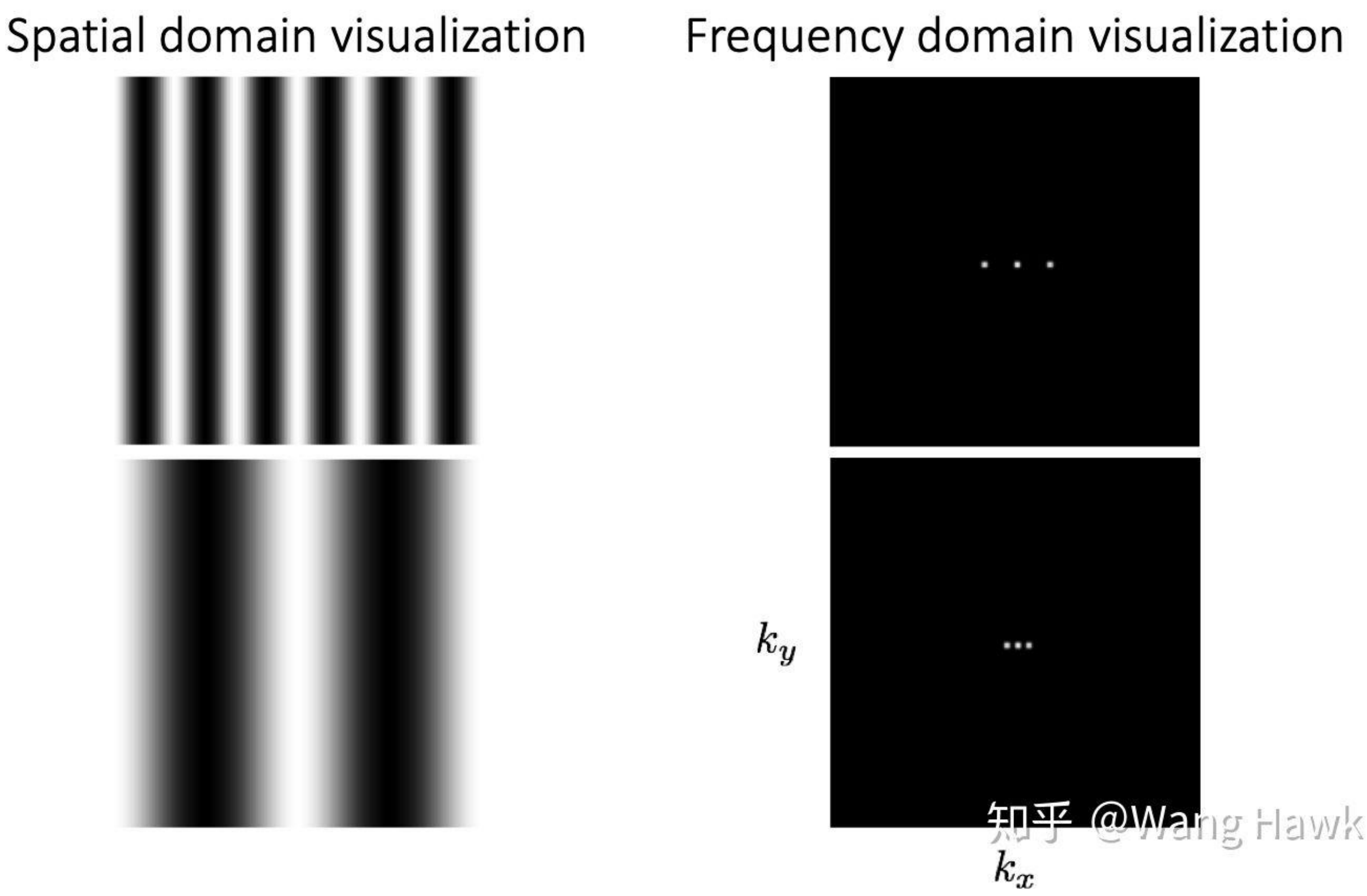
Synthesis equation

A linear combination of $\mathrm{e}^{\mathrm{j}\omega t}$

频域分析



频域分析



Spatial domain visualization
Frequency domain visualization
k_y
k_x
知乎 @Wang Hawk

卷积分析

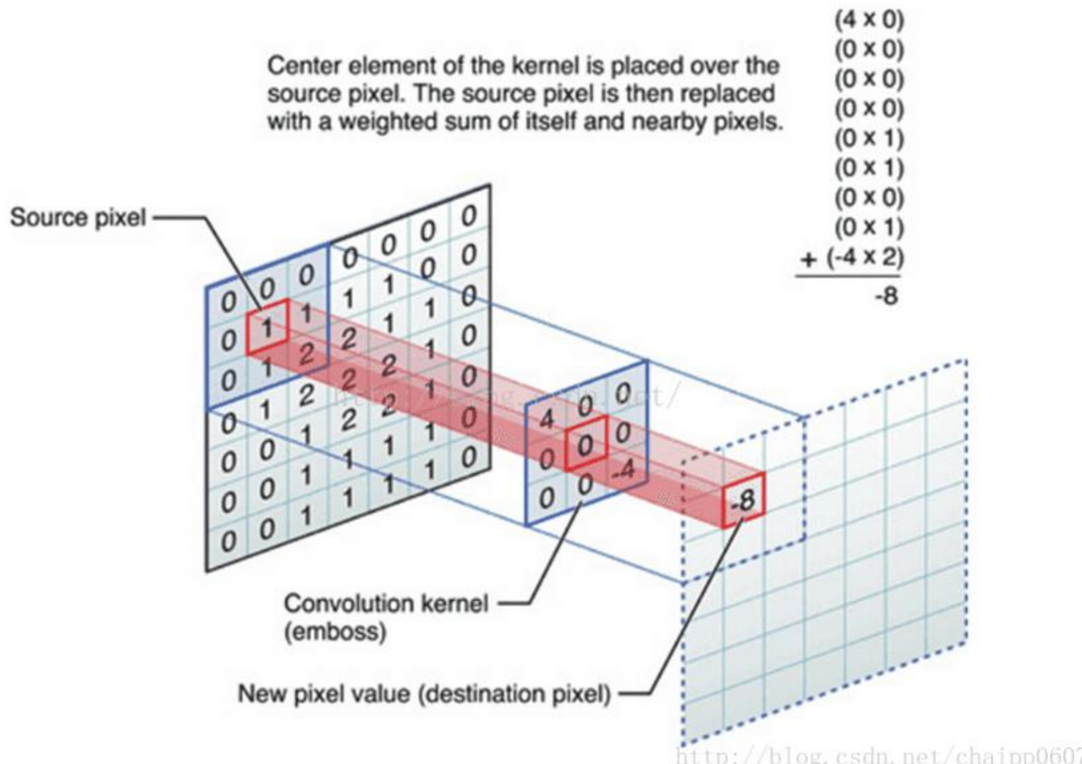
- 卷积的定义：
  - 卷积是两个变量在某范围内相乘后求和的结果。如果卷积的变量是序列x(n)和h(n)，则卷积的结果：

$$y(n)=\sum_{i=-\infty}^{\infty}x(i)h(n-i)=x(n)*h(n)$$

http://blog.csdn.net/chaipp0607

https://blog.csdn.net/chaipp0607/article/details/72236892

卷积分析

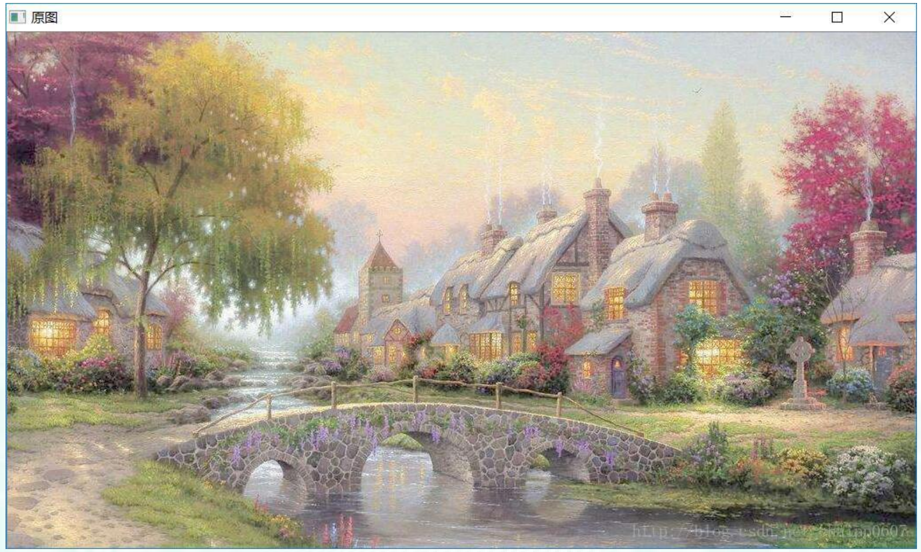


Center element of the kernel is placed over the source pixel. The source pixel is then replaced with a weighted sum of itself and nearby pixels.
(4 x 0)
(0 x 0)
(0 x 0)
(0 x 0)
(0 x 1)
(0 x 1)
(0 x 0)
(0 x 1)
+ (-4 x 2)
-8
Source pixel
Convolution kernel
(emboss)
New pixel value (destination pixel)
http://blog.csdn.net/chaipp0607

卷积分析

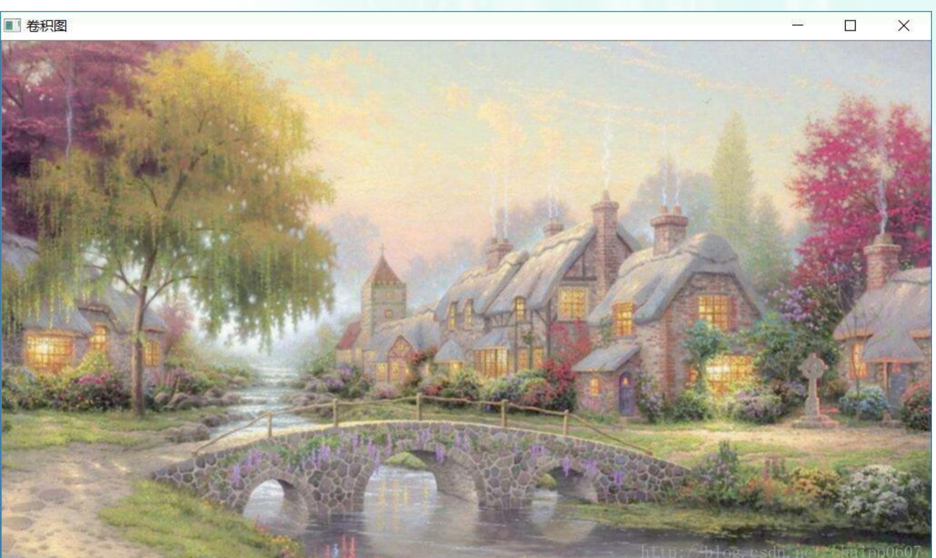
1/16	2/16	1/16
2/16	4/16	2/16
1/16	2/16	1/16

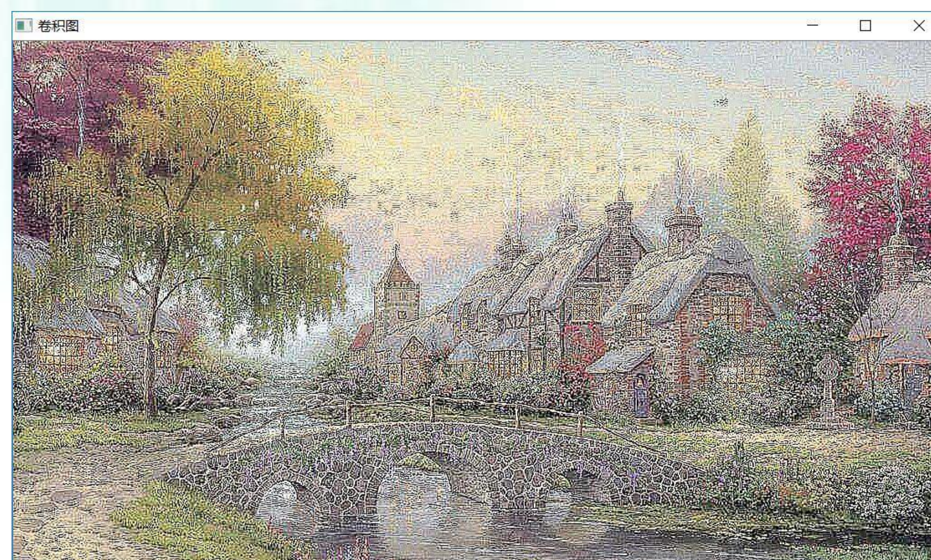
https://blog.csdn.net/chaipp0607



-1	-1	-1
-1	9	-1
-1	-1	-1

//blog.csdn.net/chaipp0607





目 录

4.4 统计信息表示方法

统计信息表示方法

- 统计直方图模型
- 词袋模型
- Skip-gram 模型

直方图表示

- 概述
  - 直方图(Histogram)，又称质量分布图，是一种统计报告图，由一系列高度不等的纵向条纹或线段表示数据分布的情况。 一般用横轴表示数据类型，纵轴表示分布情况。

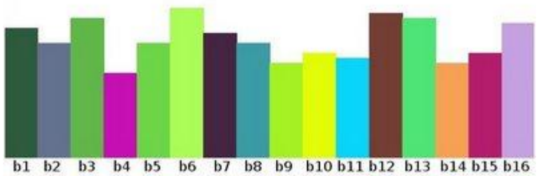
直方图表示

- 图像直方图
  - 假设有一个矩阵包含一张图像的信息（灰度值 0 – 255），既然已知数字的范围包含 256 个值，于是可以按一定规律将这个范围分割成子区域（也就是 bins）。如：

$$[0,255]=[0,15]\cup[16,31]\cup\cdots\cup[240,255]$$

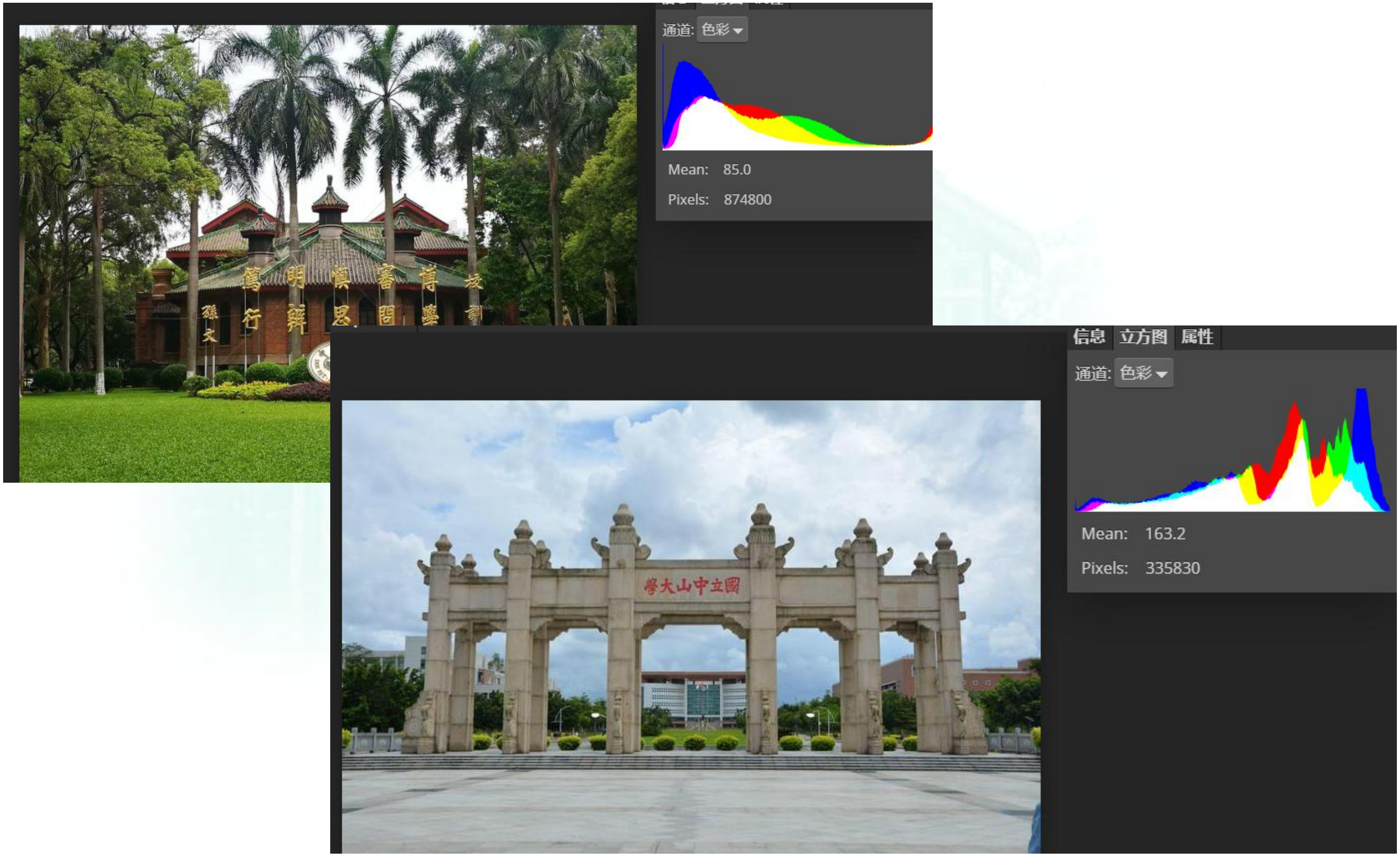
$$range=bin_1\cup bin_2\cup\cdots\cup bin_{n=15}$$

  - 然后再统计每一个 bin(i) 的像素数目。采用这一方法来统计上面的数字矩阵，可以得到下图（其中 x 轴表示 bin，y 轴表示各个 bin 中的像素个数）：



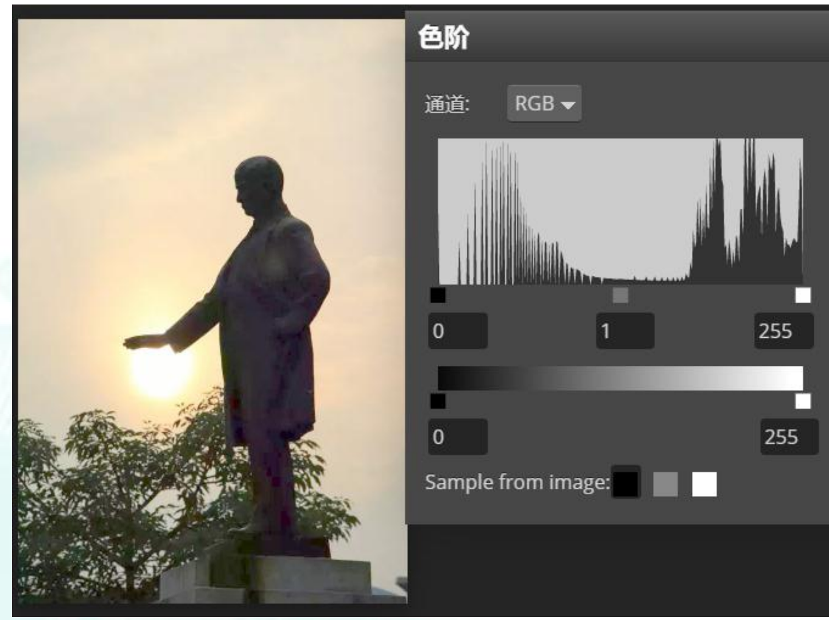
直方图表示

- 图像直方图例子



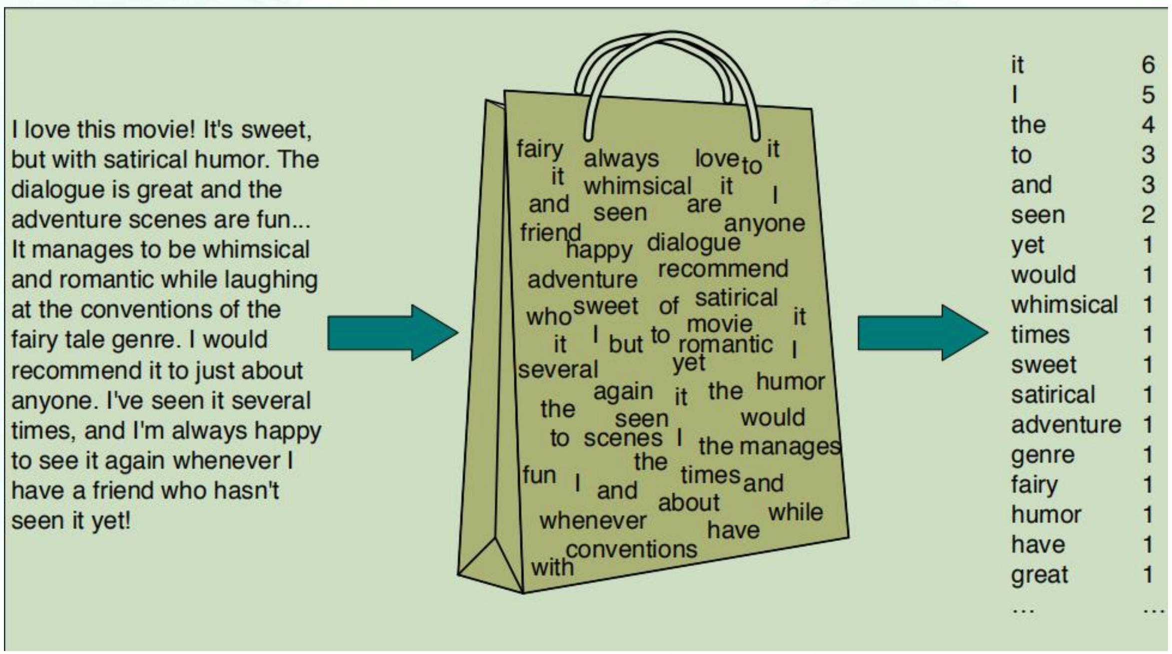
直方图表示

❑ 图像直方图例子



词袋模型

- 文本的词袋模型(Bag of Words)
  - 一种忽略顺序语法结构等信息的文本表示方法
  - 将文本序列打散，只考虑每个单词的频率。



词袋模型

- 文本的词袋模型(Bag of Words)
  - 忽略顺序的合理性？

研究表明

汉字的序顺并不定一能影阅响读

比如当你看完这句话后

仔味细回一下

才发这现话段里的字全是都乱的

词袋模型

- 图像词袋模型

词袋模型

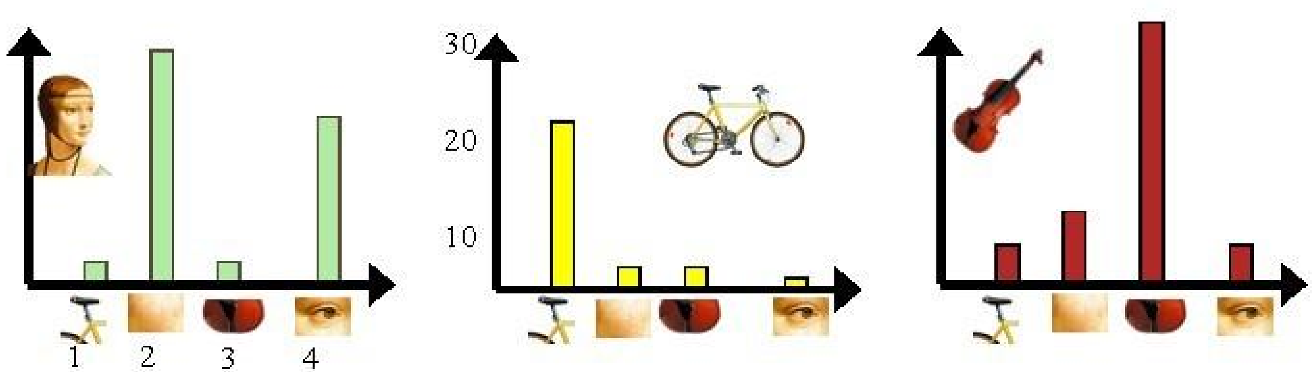
图像词袋模型

- 1）把图像分割成一个个patch，并对每个patch的中心点计算sift特征。sift算法可以提取图像中的局部不变特征，这一步是做dense sift.

词袋模型

图像词袋模型

- 2）利用kmeans算法，将这些patch的中心点的sift特征聚成为k个类，用这k个聚类中心来构造单词表。因为一幅图像中能提取出成千上万个sift向量，而每个sift向量是128维的，为了减小计算量，需要对这些sift向量做聚类，把相似的patch合并，取聚类中心作代表，构造单词表。
- 3）利用单词表表示每幅图像，则每幅图像被表示成了一个与词序列相对应的词频向量（k维的）。



Skip-gram 模型

- 语言中的顺序关系
  - 牛吃草、吃草牛、草吃牛
  - 在词袋模型中，无法表达这种顺序关系
- 统计语言模型

  对于语言序列

$$w_1, w_2, \ldots, w_n$$

  语言模型就是计算该序列的概率，即

$$P(w_1, w_2, \ldots, w_n)$$

Skip-gram 模型

- 统计语言模型
  - n-gram 语言模型
    - 引入马尔可夫假设(Markov assumption)，即假设当前词出现的概率只依赖于前 n-1 个词，可以得到

$$P(w_i|w_1,w_2,\ldots,w_{i-1}) = P(w_i|w_{i-n+1},\ldots,w_{i-1})$$

n=1 unigram: $P(w_1,w_2,\ldots,w_n) = \prod_{i=1}^{n} P(w_i)$

n=2 bigram: $P(w_1,w_2,\ldots,w_n) = \prod_{i=1}^{n} P(w_i|w_{i-1})$

n=3 trigram: $P(w_1,w_2,\ldots,w_n) = \prod_{i=1}^{n} P(w_i|w_{i-2},w_{i-1})$

Skip-gram 模型

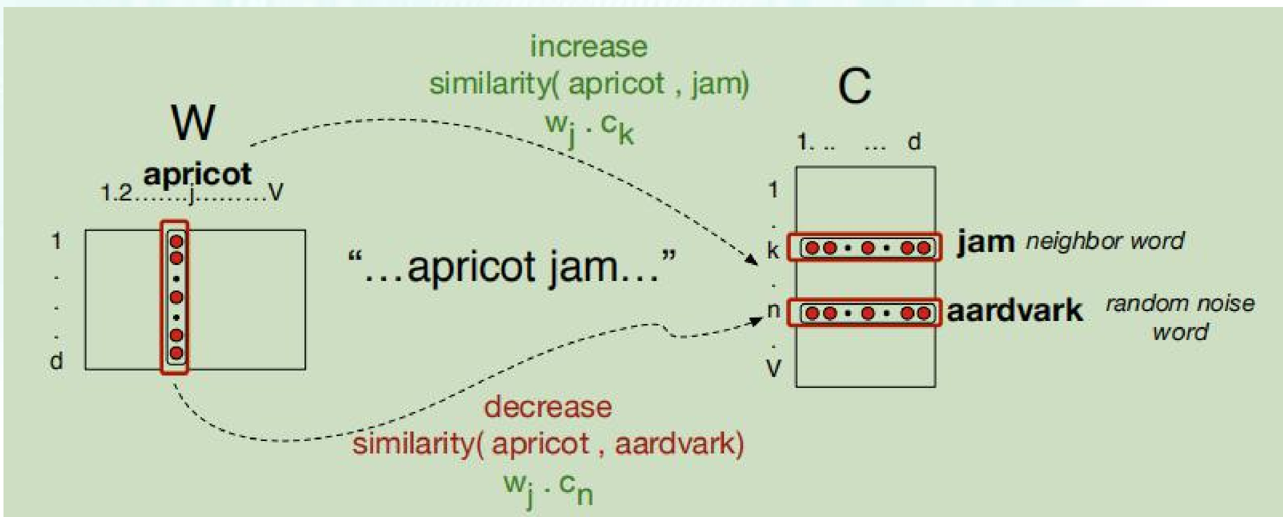
- 统计语言模型
  - skip-gram 语言模型
    - 在一个窗口内，中心词与与其上下文中出现的词的共现概率会更高。
    - 假定：选定窗口后，窗口内的上下文词的概率只与中心词有关，与其它上下文词无关。

```
... lemon,  a [tablespoon of apricot jam,      a] pinch ...
               c1          c2      t     c3        c4
```

Skip-gram 模型

- Skip-gram模型
  - 跳字模型通过选择一段窗口内的上下文，以中心词来预测窗口内的其他词，来学习如何建模词与词间的上下文关系。

```
... lemon,  a [tablespoon of apricot jam,     a] pinch ...
               c1          c2     t     c3       c4
```



参考资料

- https://zhuanlan.zhihu.com/p/52061158
- http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/
- *Speech and Language Processing*
- *Lecture: Visual Bag of Words - Stanford Vision Lab*
- https://blog.csdn.net/chaipp0607/article/details/72236892
- https://blog.csdn.net/happyer88/article/details/45769413