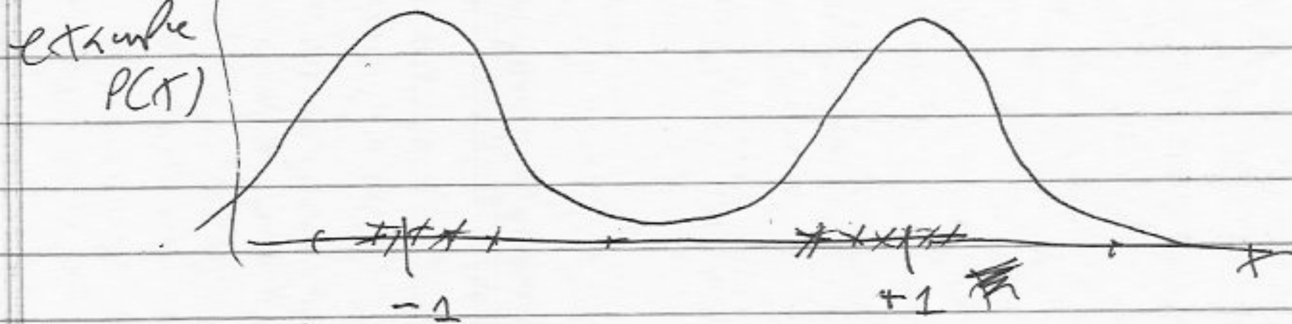
Introducing mixture of Gaussians Jan 25

Gaussian is not good at describing distributions with multiple modes (= bumps)



But could describe this as some combination of two Gaussians, $N(x|-1,\sigma^2)$ and $N(x|+1,\sigma^2)$.

alternatives.

"splicing" $P(x) \underset{\text{proportional to}}{\propto} \begin{cases} N(x|-1,\sigma^2) & x<0 \\ N(x|+1,\sigma^2) & x \geq 0 \end{cases}$

"max" ; $P(x) \propto \max\{N(x|-1,\sigma^2), N(x|+1,\sigma^2)\}$
"upper cover"

Linear combination (aka mixture of Gaussians!)

$$P(x) \overset{\text{equality!}}{=} \frac{1}{2} N(x|-1,\sigma^2) + \frac{1}{2} N(x|+1,\sigma^2)$$

Note that normalization constant of first two alternatives may be a problem to compute. But our linear combination is already correctly normalized!

$$\int_{x=-\infty}^{\infty} P(x) = \frac{1}{2}\int N(x|-1,\sigma^2) + \frac{1}{2}\int N(x|+1,\sigma^2)$$
$$= \frac{1}{2} + \frac{1}{2} = 1$$

Introducing Mixture of Gaussians

Jan 25

In general, K components with means μ_k, covariance matrices Σ_k, and "mixing weights" w_k such that

$$p(x) = \sum_{k=1}^{K} w_k N(x \mid \mu_k, \Sigma_k) \qquad \sum w_k = 1$$

again, this is easily verified to be a properly normalized density function if $\sum w_k = \underline{1}$

To generate samples from an MoG

For $i = 1$ to N

- generate U = uniform random number $U(0,1)$ between 0 and 1

- if $U < w_1$
    generate $x_i \sim N(x \mid \mu_1, \Sigma_1)$
  elseif $U < w_1 + w_2$
    generate $x_i \sim N(x \mid \mu_2, \Sigma_2)$
  elseif $U < w_1 + w_2 + \ldots + w_{k-1}$
    generate $x_i \sim N(x \mid \mu_{k-1}, \Sigma_{k-1})$
  else
    generate $x_i \sim N(x \mid \mu_k, \Sigma_k)$
  endif

endfor

Motivating EM algorithm

Jan 24, 2012

what happens when we try to do MLE (maximum likelihood estimation) of the parameters of a Gaussian mixture model?

Given N sample data points $\mathcal{X} = \{x_1, \ldots, x_N\}$ we need to estimate the mixing weights $w = \{w_1, w_2, \ldots, w_K\}$, means $\mu = \{\mu_1, \mu_2, \ldots, \mu_K\}$, and covariance matrices $\Sigma = \{\Sigma_1, \Sigma_2, \ldots, \Sigma_K\}$, of the K Gaussian components.

Assuming i.i.d samples, for the likelihood function

$$L(\mathcal{X} | w, \mu, \Sigma) = \prod_{i=1}^{N} \sum_{k=1}^{K} w_k \mathcal{N}(x_i | \mu_k, \Sigma_k)$$

Taking log likelihood gives us

$$\log L(\mathcal{X} | w, \mu, \Sigma) = \sum_{i=1}^{N} \log \sum_{k=1}^{K} w_k \mathcal{N}(x_i | \mu_k, \Sigma_k)$$

and unfortunately we are now stuck, because we don't know how to take the log of a sum. The "log" is thus prevented from getting inside the inner sum in order to work its simplification of the exponential function in $\mathcal{N}(x_i | \mu_k, \Sigma_k)$.

auxiliary derivations related to GMM-EM lecture

Jan 26
2012

$$P(x, z|\theta) = \prod_n \prod_k \left[\pi_k \, N(x_n|\mu_k, \Sigma_k)\right]^{z_{nk}}$$

$$\log P(x, z|\theta) = \sum_n \sum_k z_{nk}\left[\log \pi_k + \log N(x_n|\mu_k, \Sigma_k)\right]$$

$E_{z|x} \log P(x, z|\theta) =$ Aside: $E_{z|x}[f(z)] = f(0)P(z=0|x) + f(1)P(z=1|x)$

$$= \sum_n \sum_k 0 \cdot P(z_{nk}=0|x_n) + \left[\log \pi_k + \log N(x_n|\mu_k, \Sigma_k)\right] \cdot \underbrace{P(z_{nk}=1|x_n)}_{\text{call this } \gamma_{nk}}$$

$$= \sum_n \sum_k \gamma_{nk}\left[\log \pi_k + \log N(x_n|\mu_k, \Sigma_k)\right]$$

Notice that discrete variable $z_{nk} \in \{0, 1\}$ has been replaced with a continuous variable $0 \le \gamma_{nk} \le 1$

Now, what is $\gamma_{nk} = P(z_{nk}=1|x_n)$

Let $z_n = \{z_{n1}, z_{n2}, \dots, z_{nK}\}$ recall "1 of K" representation. only one of them is 1 and the rest are 0

$$P(x_n, z_n) = \prod_k \big[\underbrace{\pi_k N(x|\mu_k, \Sigma_k)}_{f(k)}\big]^{z_{nk}} = f(1)^{z_{n1}} f(2)^{z_{n2}} \dots f(K)^{z_{nK}}$$

$$P(z_n|x_n) = \frac{P(x_n, z_n)}{P(x_n)} = \frac{P(x_n, z_n)}{\sum_{z_n} P(x_n, z_n)}$$

Denominator $\sum_{z_n} P(x_n, z_n) = \sum_{z_n} f(1)^{z_{n1}} f(2)^{z_{n2}} \dots f(K)^{z_{nK}}$

Be careful here! This summation is over all combinations of values $z_n = \{z_{n1}, z_{n2} ..\}$ can take, which is only one of them is 1 at a time!

$$\sum_{z_n} f(1)^{z_{n1}} f(2)^{z_{n2}} \dots f(K)^{z_{nK}} = f(1) + f(2) + \dots + f(K)$$

$$\{1\,0\,0\,0 \dots,\; 0\,1\,0\,0 \dots,\; 0\,0\,1\,0 \dots\}$$

$$= \sum_j f(j) = \sum_j \pi_j N(x|\mu_j, \Sigma_j)$$

$$\boxed{\text{So } P(z_{nk}=1|x_n) = \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_j \pi_j N(x_n|\mu_j, \Sigma_j)}}$$