$p(i \mid f)$

i - номер класса

f - вектор признаков (RGB)

Правило Байеса:

$$p(i \mid f) = \frac{p(f \mid i) p(i)}{p(f)}, \; p(f) = \sum_i p(f \mid i) p(i)$$

Можно предположить что $p(f) = const$.

Если $p(i \mid f) > p(j \mid f), \forall j \neq i$, то пиксель относится к классу i. Правило Байеса может быть переписано следующим образом: если $D_i(f) > D_j(f), \forall j \neq i$. Где $D_i(f)$ - дискриминантная функция для класса:

$$D_i(f) = p(f \mid i) p(i) = p(i \mid f) p(f)$$

Свойство: решение не будет меняться при любом монотонном преобразовании функции $D_i(f)$, например:

$$D_i(f) = \ln\left[p(f \mid i) p(i) \right]$$

Для метода максимального правдоподобия k-мерную функцию $D_i(f)$ можно записать как:

$$D_i(f) = \ln\left[p(i) \right] - \frac{1}{2}\left[k \ln(2\pi) + \ln\left\| \Sigma_i \right\| + (f - \mu_i)^T \Sigma_i^{-1} (f - \mu_i) \right]$$

Предположим что $p(i) = const$:

$$D_i(f) = -\ln\left\| \Sigma_i \right\| - (f - \mu_i)^T \Sigma_i^{-1} (f - \mu_i)$$

Отбросим $-\ln\left\| \Sigma_i \right\|$ и получим метод расстояния Махаланобиса:

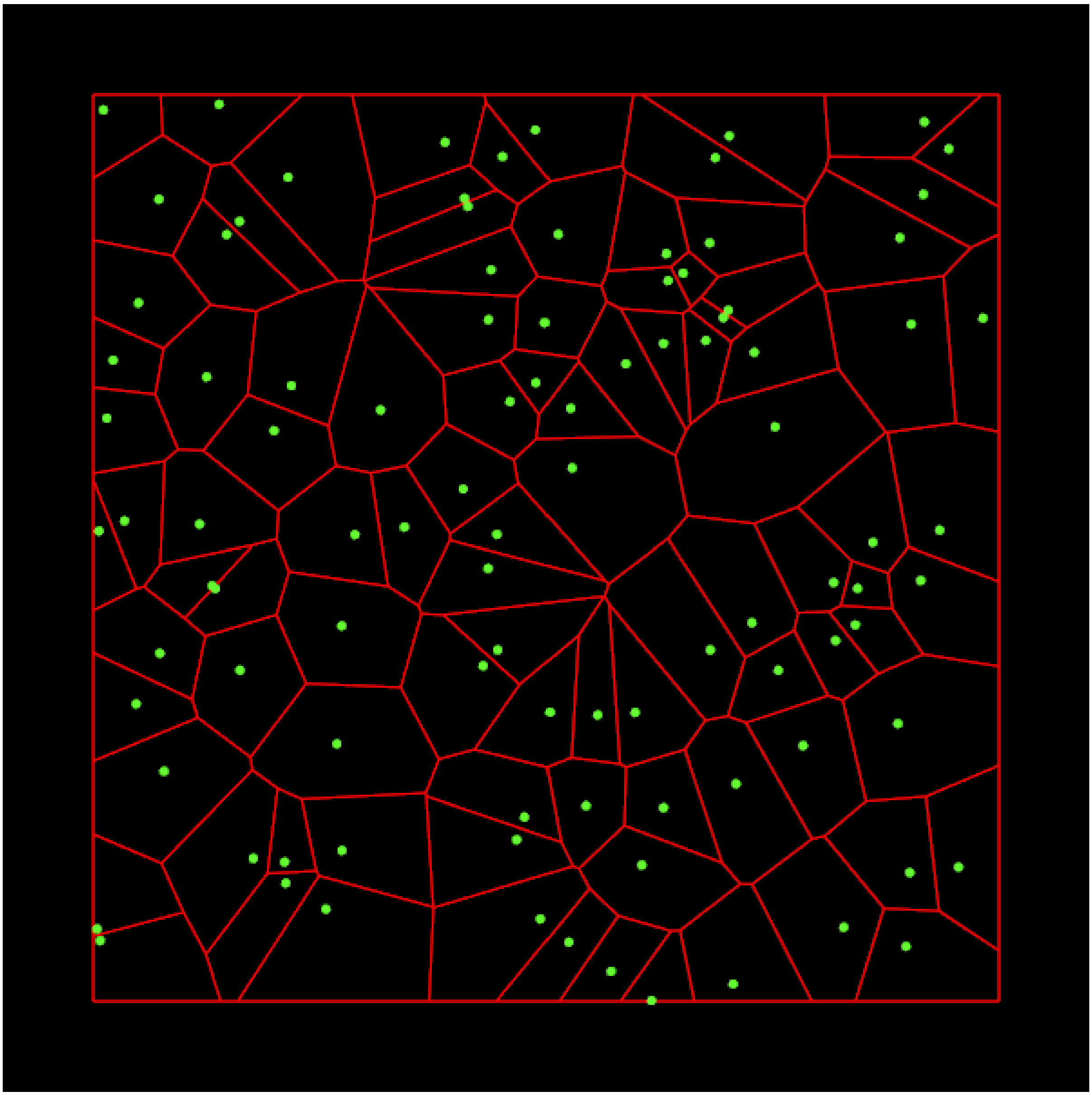
$$D_i(f) = -(f - \mu_i)^T \Sigma_i^{-1} (f - \mu_i)$$

Предположим, что

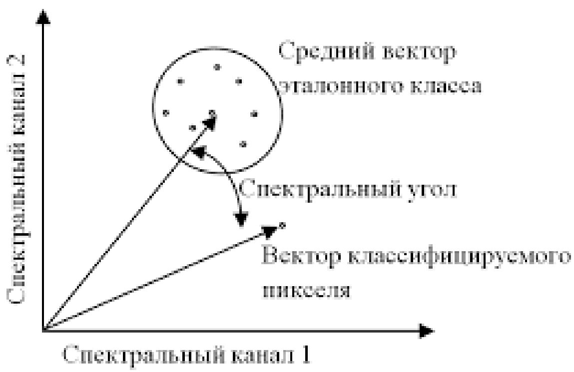
$$\Sigma_i = \begin{pmatrix} c & 0 & \dots & 0 \\ 0 & c & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & c \end{pmatrix}$$

и получим метод минимального расстояния:

$$D_i(f) = -(f - \mu_i)^T (f - \mu_i) = -\operatorname{dist}(f, \mu_i)$$



Метод спектрального угла: где меньше угол такой и класс.



$$D_i(f) = \cos[f \wedge \mu_i] = \frac{(f, \mu_i)}{\|f\|\|\mu_i\|} \sim \left(f, \frac{\mu}{\|\mu_i\|} \right)$$

Метод k-средних

Минимизируется среднеквадратическое отклонение от центров кластеров:

$$V = \sum_{i=1}^{k} \sum_{f \in S_i} \mathrm{dist}(f, \mu_i)^2 \to \min$$

Алгоритм:

$\mu_i^{(0)}$ - задается

1. Выполняется распределение всех пикселей по кластерам S_i согласно:

$$D_i(f) = -\mathrm{dist}(f, \mu_i^{(n)})$$

2. Обновляются центры кластеров $\mu_i^{(n+1)} = \dfrac{1}{|S_i|} \sum_{f \in S_i} f$

Условие завершения: $\mu_i^{(n)} = \mu_i^{(n+1)}$

Недостатки:

1. Не гарантируется достижение глобального минимума суммарного квадратичного отклонения V, а только одного из локальных минимумов.
2. Результат зависит от выбора исходных центров кластеров, их оптимальный выбор неизвестен.
3. Число кластеров надо знать заранее.

Решение: алгоритм ISODATA