# Normal Classification Model [5]

Qi Zhao

July 30, 2018

## 1. Introduction

As a representative problem when considering face detection [4]; it observes a 60 60 RGB [3] image patch and whether it contains a face or not. To this end, it concatenates the RGB values to form the 108001 vector x. The goal is to take the vector x and return a label w $\in$ 0,1 indicating whether it contains background (w = 0) or a face (w = 1). In a real face detection system it would repeat this procedure for every possible sub-window of an image (Figure 1). It will start with a basic generative approach in which it describes the likelihood of the data in the presence/absence of a face with a normal distribution [1].

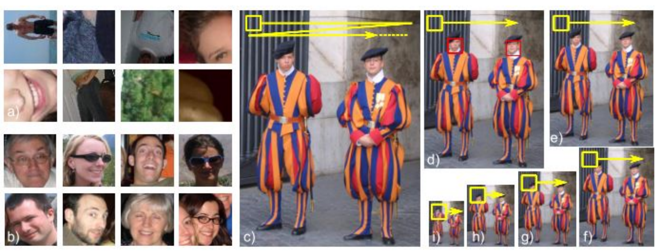


Figure 1. Face detection. Consider examining a small window of the image (here 60 60). We concatenate the RGB values in the window to make a data vector x of dimension 10800 1. The goal of face detection is to infer a label w $\in$ {0,1} indicating whether the window contains (a) a background region (w = 0) or (b) an aligned face (w = 1). (c-i) We repeat this operation at every position and scale in the image by sweeping a fixed size window through a stack of resized images, estimating w at every point.

It will take a generative approach to face detection; It will model the probability of the data x and parameterize this by the world state w. It will describe the data with a multivariate normal distribution in Equation 1

$$Pr(x|w) = Norm_x[\mu_w, \Sigma_w] \tag{1}$$

or treating the two possible values of the state w separately, it can explicitly write in Equation 2.

$$\begin{aligned} Pr(x|w=0) &= Norm_x[\mu_0, \Sigma_0] \\ Pr(x|w=1) &= Norm_x[\mu_1, \Sigma_1] \end{aligned} \tag{2}$$

These expressions are examples of class conditional density functions. They describe the density of the data x conditional on the value of the world state w.

## 2. Descriptions

The goal of learning is to estimate the parameters $\theta$ = $\mu_0$, $\Sigma_0$, $\mu_1$, $\Sigma_1$ from example pairs of training data $\{x_i, w_i\}_{i=1}^{I}$. Since parameters $\mu_0$ and $\Sigma_0$ are concerned exclusively with background regions (where w=0) we can learn them from the subset of training data $S_0$ that belonged to the background.

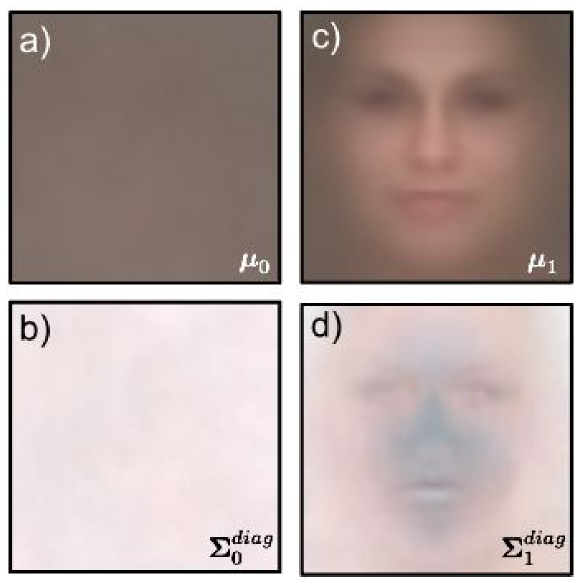


Figure 2. Class conditional density functions for normal model with diagonal covariance. Maximum likelihood fits based on 1000 training examples per class. a) Mean for background data $\mu_0$ (reshaped from 108001 vector to 60 60 RGB image). b) Reshaped square root of diagonal covariance for background data $\Sigma_0$. c) Mean for face data $\mu_1$. d) Covariance for face data $\Sigma_1$. The background model has little structure: the mean is uniform and the variance is high everywhere. The mean of the face model clearly captures class-specific information. The covariance of the face is larger at the edges of the image, which usually contain hair or background.

Similarly, $\mu_1$ and $\Sigma_1$ are concerned exclusively with faces (where w = 1) and can be learned from the subset $S_1$ of training data which contained faces. Figure 2 shows the maximum likelihood estimates of the parameters where it has used the diagonal form of the covariance matrix [2].

The goal of the inference algorithm is to take a new facial image x and assign a label w to it. To this end, we define a prior over the values of the world state Pr(w) = Ber$n_w$ [] and apply Bayes rule [6] in Equation 3.

$$Pr(w = 1|x) = \frac{Pr(x|w = 1)Pr(w = 1)}{\sum_{i=1}^{1} Pr(x|w = k)Pr(w = k)} \tag{3}$$

## References


[1] A. Azzalini and A. D. Valle. The multivariate skew-normal distribution. *Biometrika*, 83(4):715–726, 1996. 1

[2] D. E. Gustafson and W. C. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *IEEE Conference on Decision and Control*, pages 761–766, 1979. 2

[3] K. Lai, L. Bo, X. Ren, and D. Fox. A large-scale hierarchical multi-view RGB-D object dataset. In *IEEE International Conference on Robotics and Automation*, pages 1817–1824, 2011. 1

[4] E. Osuna, R. Freund, and F. Girosit. Training support vector machines: an application to face detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 130–136, 1997. 1

[5] M. Panjehpour, B. F. Overholt, J. L. Schmidhammer, C. Farris, P. F. Buckley, and T. Vo-Dinh. Spectroscopic diagnosis of esophageal cancer: new classification model, improved measurement system. *Gastrointestinal Endoscopy*, 41(6):577–581, 1995. 1

[6] R. A. Waller and D. B. Duncan. A Bayes' rule for the symmetric multiple comparisons problem. *Journal of the American Statistical Association*, 64(328):1484–1503, 1969. 2