1. Project is about inferring if the given picture composed of a face or a noise (background). Here we use different classification methods to classify if a test image belongs to face or noise categories.
2. Basically, the approach is fundamentally a logistic classification in principle with different coatings to improve the performance. Here, the theme is to represent input face classification as a pdf. Further, this pdf curve is made to be parameterized on the input data. So, the output (label classification) is characterized as a function of the input images with corresponding weights as coefficients to these images. In the learning stage, these weights are learnt. However, the Ml approach has a challenge in that the problem is not a closed-form and hence iterative optimization process is employed to determine the phi vector. The classification of the test images in inferred using these phi values.

In case of Bayesian Logistic regression, a prior probability distribution for the phi is considered and laplace approximation is employed to represent the conditional probability of phi given X and phi, with a normal pdf. After learning the phi values, the inference is made possible.

In dual logistic regression, phi is factorized into a product of X and psi. Where psi has a dimension of the length of the input images whereas phi is of length of pixels in each input image. The advantage of using psi is significant when the input image dimensions are too big and computationally expensive.

In Dual Bayesian logistic regression, we combine both the prior and psi approach to determine the coefficients for inference on test images.

Kernel logistic regression involves computing a dot product of input image matrix with itself. So, the product is often delegated to an operation called kernel to compute it. Often called Kernel trick. Many ways to compute kernel, in the assignment RBF is considered. So, the delegation has a rippling effect causing changes in the equations slightly. Otherwise pretty much the same.

In Relevance Vector Classification we intend to work on sparse data. For this we impose a penalty on non-zeros weight images. After that we represent the normal prior with a product of one dimensional t-distributions. When we rewrite each t-distribution as a marginalization of the join distribution and doing subsequent approximations we get the psi values through which we infer the classification of the test images into belonging to face or nonface.

1. Results

|  |  |  |
| --- | --- | --- |
| Method | Miss detection | False alarm |
| Logistic Regression | 0.055 | 0.017 |
| Bayesian Logistic Reg | 0.057 | 0.054 |
| Dual Logistic | 0.036 | 0.035 |
| Dual Bayesian Logistic | 0.036 | 0.035 |
| Kernel Logistic Reg | 0.118 | 0.081 |
| Relevance Vector Log Reg | 0.163 | 0.002 |

Supposedly, the performance of the dual and Bayesian logistic regression to be better than the logistic regression but due to the unimaginative initial values chosen for the involved parameters explains the astray of the performance of the methods shown in the above table. It can be noticed from the above table the initialization of parameters hold sway as to how the respective models perform and hence a detailed study of the performance variance under different settings is recommended. In my undertaking though on contrary I didn’t find the difference in performance especially noteworthy.