Exponential Family Discriminant Analysis: Generalizing LDA-Style Generative Classification to Non-Gaussian Models

Anish Lakkapragada
Yale University
New Haven, CT 06511
anish.lakkapragada@yale.edu

Abstract

We introduce Exponential Family Discriminant Analysis (EFDA), a unified framework that extends classical Linear Discriminant Analysis (LDA) generative classification beyond the Gaussian setting to any member of the exponential family. Under the assumption that each class–conditional density belongs to the same exponential family, EFDA derives closed-form or semi-closed-form maximum likelihood estimators for all natural parameters, and yields an explicit expression for the log-odds ratio as an linear function of the sufficient statistic. We demonstrate the method on the Weibull distribution, showing that EFDA accurately models the nonlinear log-odds ratio, which logistic regression and LDA are unable to capture. Finally, to demonstrate EFDA's practicality, we provide closed-form EFDA derivations for four additional exponential family distributions.

1 Introduction

Discriminant analysis is a cornerstone of generative classification, dating back to Fisher's Linear Discriminant Analysis (LDA) which assumes that each class–conditional distribution is Gaussian with a shared covariance matrix. While LDA offers closed-form solutions and interpretable parameters, its reliance on normality limits applicability to non-Gaussian data. Logistic regression, a discriminative alternative, models the log-odds directly but requires iterative fitting and is constrained to only linear functions [2].

Exponential families encompass a broad class of distributions (e.g. Normal, Gamma, Poisson, Weibull, Bernoulli) that admit the density form stated below:

$$f(\mathbf{x} \mid \eta) = h(\mathbf{x}) \exp(\eta \cdot T(\mathbf{x}) - A(\eta))$$

where $T(x)$ is the sufficient statistic and η the natural parameter. While generative classifiers built on exponential families include Naive Bayes and Generalized Linear Models ([1, 4]), there has been little work on directly extending LDA's technique of modeling class-conditional densities for generative classification in non-Gaussian scenarios. To the best of our knowledge, studied kernel-based approaches such as [3] and [6] focus on dimensionality reduction rather than generative classification.

In this paper we develop *Exponential Family Discriminant Analysis (EFDA)*, which:

1. Assumes each class–conditional distribution belongs to the same exponential family.
2. Derives maximum likelihood estimators for the class priors and natural parameters η_0, η_1 via solvable equations involving $T(x)$.

3. Models the log-odds ratio as

$$\ell(x) \;=\; \log \frac{P[Y=1 \mid x]}{P[Y=0 \mid x]} \;=\; \log \frac{\alpha}{1-\alpha} + \big[A(\eta_0) - A(\eta_1)\big] + (\eta_1 - \eta_0) \cdot T(x),$$

resulting in a decision boundary that is linear in $T(x)$.

We validate EFDA in simulation using the Weibull distribution, comparing its estimated log-odds curves against those from logistic regression. Our results highlight EFDA's superior ability to capture nonlinear boundaries when the true generative model is exponential-family–structured. We conclude by deriving closed-form EFDA solutions for four other common distributions and providing directions for future work.

2 Background

2.1 Linear Discriminant Analysis

We establish some preliminaries: supppose we have covariates $X = (x_1, \ldots, x_p) \in \mathbb{R}^p$ and binary response $Y \in \{0, 1\}$. Our end goal is to model $p(X) = \mathbb{P}[Y = 1 \mid X]$, as this allows us to construct the Bayes classifier h (defined below):

$$h(x) = \begin{cases} 1 & \text{if } p(x) \geq 0.5 \\ 0 & \text{else} \end{cases} \tag{1}$$

Linear Discriminant Analysis (LDA) is one such method to approach this binary classification problem. To do so, it learns the class-conditional densities $X \mid Y$ so as to create a (Bayes) classifier that models $Y \mid X$. Let us define $f_0(x)$ and $f_1(x)$ as the class-conditional densities of $X \mid Y = 0$ and $X \mid Y = 1$ respectively, and $\alpha = \mathbb{P}[Y = 1]$, or the marginal probability of $Y = 1$ across all data. LDA then gives the following form for $p(X)$:

$$p(X) = \frac{\mathbb{P}[X \mid Y=1]\mathbb{P}[Y=1]}{\mathbb{P}[X \mid Y=1]\mathbb{P}(Y=1) + \mathbb{P}[X \mid Y=0]\mathbb{P}(Y=0)} = \frac{\alpha f_1(x)}{\alpha f_1(x) + (1-\alpha) f_0(x)} \tag{2}$$

Thus, to model $p(X)$, LDA must learn $f_0(x)$, $f_1(x)$, and α. To do so, it assumes that $X \in \mathbb{R}^p$ has a normal distribution for each class with a shared covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$. Restated, LDA posits:

$$X \mid Y = 1 \sim \mathcal{N}(\mu_{\mathbf{1}}, \boldsymbol{\Sigma}), \quad X \mid Y = 0 \sim \mathcal{N}(\mu_{\mathbf{0}}, \boldsymbol{\Sigma})$$

with shared covariance matrix $\boldsymbol{\Sigma}$ across both classes. Then parameters $\mu_{\mathbf{1}}, \mu_{\mathbf{0}}, \boldsymbol{\Sigma}$ (and α) are learned through MLE on the given dataset $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$. For class $k \in \{0, 1\}$, the MLE estimates are:

$$\hat{\alpha} = \frac{N_1}{n}, \quad \hat{\mu}_k = \frac{1}{N_k} \sum_{i:Y_i=k} X_i, \quad \hat{\Sigma} = \frac{1}{n} \sum_{k=0}^{1} \sum_{i:Y_i=k} (X_i - \hat{\mu}_k)(X_i - \hat{\mu}_k)^T.$$

where N_k is the number of observations in $\mathcal{D}$ belonging to class k.

We conclude this section with providing some understanding of the commonalities and differences between logistic regression and LDA. We start by giving the log-odds ratio of the LDA:

$$\log \frac{\mathbb{P}[Y=1 \mid X]}{\mathbb{P}[Y=0 \mid X]} = \log \frac{\alpha f_1(x)}{(1-\alpha) f_0(x)} = \log \frac{\alpha}{1-\alpha} + \log \frac{\exp(-\frac{1}{2}(x-\mu_{\mathbf{1}})^T \Sigma^{-1}(x-\mu_{\mathbf{1}}))}{\exp(-\frac{1}{2}(x-\mu_{\mathbf{0}})^T \Sigma^{-1}(x-\mu_{\mathbf{0}}))}$$

$$= \underbrace{\log \frac{\alpha}{1-\alpha} - \frac{1}{2}\mu_{\mathbf{1}}{}^T \boldsymbol{\Sigma}^{-1} \mu_{\mathbf{1}} + \mu_{\mathbf{0}}{}^T \boldsymbol{\Sigma}^{-1} \mu_{\mathbf{0}}}_{\beta_0} + \sum_{i=1}^{p} \underbrace{[\boldsymbol{\Sigma}^{-1}(\mu_{\mathbf{1}} - \mu_{\mathbf{0}})]_i}_{\beta_i} \cdot X_i$$

We can see that the log-odds ratios in LDA are modeled as a linear function of x (hence the name *Linear* Discriminant Analysis). Furthermore, these coefficients in LDA can be thought of as parametrized versions of the logistic regression parameters β. In short, LDA provides a clean closed-form solution and good interpretability, in exchange for a strong parametric assumption.

2.2 Exponential Family Distributions

In this paper, we present a generalization of LDA to data distributed by exponential families. As such, in this section we provide an abbreviated introduction to the exponential family in statistics.

An *exponential family* is a set of probability distributions whose PDF $f(\mathbf{x} \mid \eta)$ can be expressed as:

$$f(\mathbf{x} \mid \eta) = h(\mathbf{x}) \exp(\eta \cdot T(\mathbf{x}) - A(\eta)) \tag{3}$$

Before explaining the meaning of these functions in the above representation of our PDF $f(\mathbf{x} \mid \eta)$, it is important to note that many common distributions that we encounter fall into the exponential family (e.g. Normal, Poisson, Gamma, Weibull).

We now discuss two of these functions in our exponential family parametrized density:

1. $T(\mathbf{x}) \in \mathbb{R}^d$ (for some $d \in \mathbb{N}$) is the sufficient statistic [5]. This means that $T(\mathbf{x})$ holds all possible information about parameter θ. Alternatively expressed, conditional distribution $\mathbf{x} \mid T(\mathbf{x})$ does not depend on θ.
2. $\eta \in \mathbb{R}^d$ is the natural parameter [5]. For our purposes, it is a convenient re-expression of θ amenable to this form (meaning that we have some (ideally invertible) function $\theta \mapsto \eta$).

The remaining functions $A(\eta) \in \mathbb{R}$ and $h(\mathbf{x}) \in \mathbb{R}$ are not too important for our purposes other than to ensure that $f(\mathbf{x} \mid \eta)$ is a valid PDF (the former can be thought of as a normalization constant).

3 Exponential Family Discriminant Analysis

Having established all background requisite knowledge, we are ready to understand how we can generalize LDA to data with class-conditional distributions in the exponential family. With a slight abuse of terminology[1], let us call this new procedure we are outlining as exponential family discriminant analysis (EFDA).

We start with our assumptions. We assume we have two classes of data where data in class one and zero are distributed according to respective exponential family PDFs $f_1(\mathbf{x} \mid \eta_1)$ and $f_0(\mathbf{x} \mid \eta_0)$ such that:

$$\underset{k \in \{0,1\}}{f_k(\mathbf{x} \mid \eta_k)} = h(\mathbf{x}) \exp[\eta_k \cdot T(\mathbf{x}) - A(\eta_k)]$$

First note that PDFs f_1 and f_0 only differ by their true (natural) parameter values η_1 and η_0. Furthermore, we will model $p(X)$ in the same way as LDA – by learning the parameters present in (2). Specifically, we will estimate through MLE parameters α, η_1, and η_0. We now derive these MLEs. We start with our log-likelihood function:

$$\mathcal{L}(\alpha, \eta_0, \eta_1) = \sum_{i=1}^{n} \mathbf{1}\{Y_i = 1\} \log[\mathbb{P}(X_i \mid Y_i = 1)\mathbb{P}(Y_i = 1)] + \mathbf{1}\{Y_i = 0\} \log[\mathbb{P}(X_i \mid Y_i = 0)\mathbb{P}(Y_i = 0)]$$

$$= \sum_{i=1}^{n} \log[h(X_i)] + \mathbf{1}\{Y_i = 1\}[\log(\alpha) + \eta_1 \cdot T(X_i) - A(\eta_1)] + \mathbf{1}\{Y_i = 0\}[\log(1 - \alpha) + \eta_0 \cdot T(X_i) - A(\eta_0)]$$

[1]The reason for why this is technically misleading is that the terminology for L/QDA is based on how the log-odds function is related to X. As we will see, this can differ based on the type of distribution our data takes (regardless of it is in the exponential family.)

With some simple calculus (see Appendix), we find that the MLEs for η_1 and η_0 do not have a general closed-form solution. This is because the function $A(\eta)$ will differ among distributions in the exponential family. For example, $A(\eta) = e^{\eta}$ for the Poisson distribution but $A(\eta) = \log(1 + e^{\eta})$ for the Bernoulli distribution. As such, the best we can do is give a a solvable equation for the MLEs, which we present below:

$$\boxed{\hat{\alpha} = \frac{N_1}{n}, \quad \sum_{i:Y_i=1}^{n} T(X_i) = N_1 \cdot \frac{\partial A(\eta_1)}{\partial \eta_1}, \quad \sum_{i:Y_i=0}^{n} T(X_i) = N_0 \cdot \frac{\partial A(\eta_0)}{\partial \eta_0}} \tag{4}$$

To resummarize, we have now stated the general procedure for performing EFDA, regardless of which exponential family distribution we assume our class-conditional data densities are distributed by. Before testing EFDA, we provide one more fact: under EFDA assumptions, the true log-odds ratio $\ell(X)$ as a function of X has a clean form. See below.

$$\ell(X) = \log[\frac{\alpha f_1(X)}{(1-\alpha) f_0(X)}] = \log[\frac{\alpha}{(1-\alpha)}] + [A(\eta_0) - A(\eta_1) + T(x) \cdot (\eta_1 - \eta_0)] \tag{5}$$

and so we can see that under our EFDA assumptions, the true log-odds ratio $\ell(X)$ is a linear function of sufficient statistic $T(x)$. As we will see, this sufficient statistic $T(x)$ is frequently not a linear function of X.

We now test EFDA.

4 Experiments & Further Derivations

4.1 Worked Example: Weibull Distribution

We first show the efficacy of EFDA by putting it to the test in a simulation. For this simulation, let us assume that our data in each class is distributed according to the Weibull distribution, which does have an exponential family parametrization. We can express this distribution as Weibull(λ, k), where parameters $\lambda > 0$ (scale) and $k > 0$ (shape). Moreover, the Weibull PDF can be given by:

$$f(x) = \begin{cases} \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} \exp(-(\frac{x}{\lambda})^k) & \text{if } x \geq 0 \\ 0 & \text{else} \end{cases}$$

We explicitly state our assumptions for our data in each class:

$$X \mid Y = 0 \sim \text{Weibull}(\lambda_1, k) \quad X \mid Y = 1 \sim \text{Weibull}(\lambda_0, k)$$

Recall that our goal with EFDA is to be able to learn all required parameters to be able to model classifier function $p(X) = \mathbb{P}[Y = 1 \mid X]$. For this simulation, we specify the following ground truth values:

$$k = 3 \quad \lambda_1 = 2 \quad \lambda_0 = 4 \quad \alpha = 0.7$$

Note that $\alpha = 0.7$ means that there is an overall bias towards data belonging in class one. To help understand the true distribution of our data, we provide Figure 1.

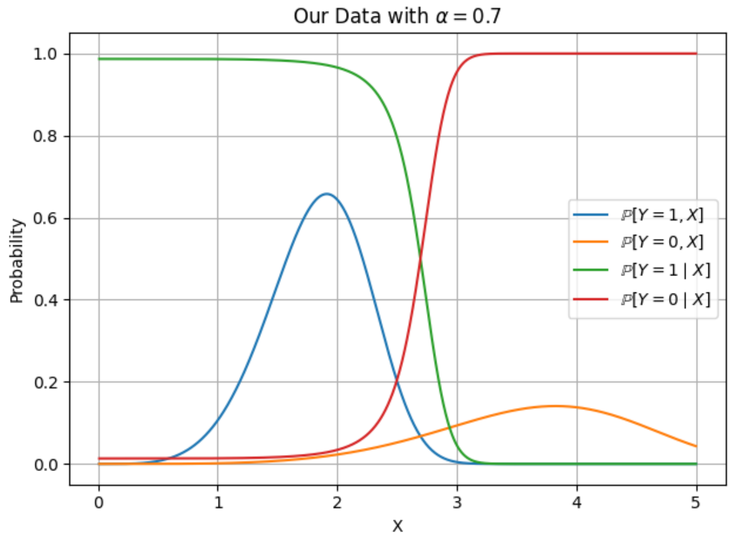

Figure 1: The most important curve in the figure is the green curve, which shows $p(X)$. As we can see, the probability of data belonging in class one sharply decreases in the range of x from 2 to 3.

Using (5), we also give the log-odds ratio $\ell(X)$ for our data:

$$\ell(X) = \log[\frac{\alpha}{(1-\alpha)}] + [A(\eta_0) - A(\eta_1) + T(x) \cdot (\eta_1 - \eta_0)] = \log\frac{\alpha}{1-\alpha} + k\log\frac{\lambda_0}{\lambda_1} + x^k(\frac{1}{\lambda_0^k} - \frac{1}{\lambda_1^k})$$

We now proceed to actually use our derived EFDA procedure to model $p(X)$ from our given dataset $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$. Note that for our simulation we are assuming $k = 3$ to be known a priori. As a first step to start deriving our EFDA procedure, we give the exponential family parametrization of the Weibull PDF, for when k is known a priori[2]:

$$\eta = -\frac{1}{\lambda^k}, \quad h(x) = x^{k-1}, \quad T(x) = x^k, \quad A(\eta) = \log(-\frac{1}{\eta k})$$

Recall that we are assuming that the true natural parameters η_1 and η_0 are different across each class. Now applying the derived MLE conditions stated in (4) for this specific exponential family parametrization, we can find the MLEs of η_1 and η_0. We start by deriving the former:

$$\sum_{i:Y_i=1}^{n} T(X_i) = N_1 \cdot \frac{\partial A(\eta_1)}{\partial \eta_1} \implies \sum_{i:Y_i=1}^{n} X_i^k = N_1 \cdot (-\frac{1}{\eta_1}) \implies \hat{\eta}_1 = -\frac{N_1}{\sum_{i:Y_i=1}^{n} X_i^k}$$

and so by identical logic we have:

$$\hat{\eta}_0 = -\frac{N_0}{\sum_{i:Y_i=0}^{n} X_i^k}$$

As a point of reference, we benchmark EFDA against standard logistic regression. Because the two methods posit inherently different parameterizations, we do not attempt a term-by-term comparison of their estimates. Instead, for each method we examine its estimated log-odds function as a function of x. To assess both bias and variability, we repeat the following procedure $M = 100$ times:

[2]For a verification this leads to the desired result, see Appendix.

1. Sample $n = 10^4$ i.i.d. observations[3] (of which roughly 70% will be in class one).
2. Fit EFDA by maximum likelihood and, separately, fit a logistic regression model.
3. Compute each method's estimated log-odds curve over a fine grid of x-values.

Figure 2 plots the resulting M log-odds curves for EFDA and for logistic regression.

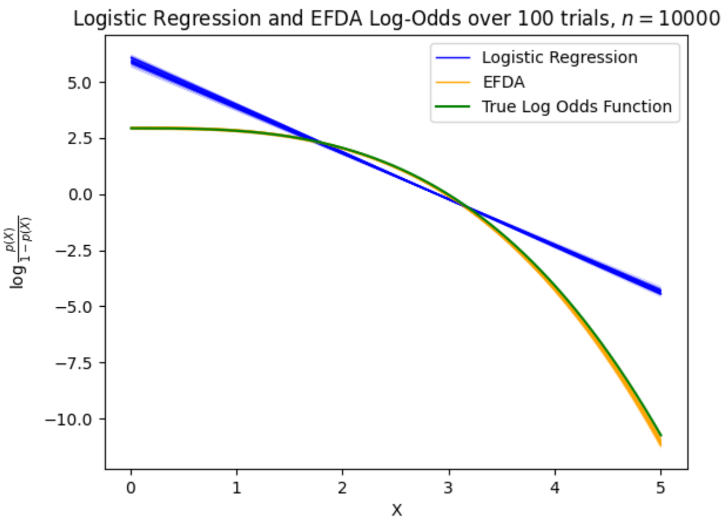

Figure 2: Plot of log-odds curves for EFDA & Logistic Regression across 100 different trials, each with a sample size of $n = 10^4$. We also provide a plot of the true log-odds function $\ell(X)$ as per our predecided ground-truth parameters for this simulation.

The major takeaway from this plot is that in cases where the true log-odds function is nonlinear (green curve), Logistic Regression will fail to model this nuance whereas a correctly-specified EFDA classifier will be able to. Furthermore, because Logistic Regression (and LDA) can only model the log-odds linearly, this means that for certain values of x (e.g. small x), it will overconfidently predict one class, and for other values of x (e.g. large x), it will underconfidently predict the other class.

4.2 EFDA Derivations for other Exponential Family Distributions

EFDA, as we saw in the simulation, comes with a standard procedure. Namely, based on our assumption of our data's class-conditional densities, we use the exponential family parametrization for $\eta, A(\eta), T(x)$ to derive the MLEs[4] of η_1 and η_0 by solving the conditions stated in (4).

To further demonstrate EFDA's widespread applicability, we derive these MLEs for a handful of distributions belonging to the exponential family. In Table 1, for each distribution, we provide:

1. The required $A(\eta)$ and $T(x)$ in the distribution's exponential family parameterization.
2. The MLEs $\hat{\eta}_1$ and $\hat{\eta}_0$ for this distribution (represented as $\hat{\eta}_k$ for class $k \in \{0, 1\}$)

From Table 1, we can see that for a wide spectrum of common distributions belonging in the exponential family, EFDA has a closed-form solution which is easy to compute. This is contrast to Logistic Regression, which is always learned through some form of iterative optimization.

[3] We provide a brief explanation on how we sample a single observation for this simulation. To sample a single observation, we follow these two steps: (1) choose whether the observation will be in class one or zero, with the former having a 70% bias (2) sample from that class's class-conditional data distribution (i.e. distribution $X \mid Y = k$ for class $k \in \{0, 1\}$).

[4] Note that MLE $\hat{\alpha}$ is same across all distributions.

Distribution	$A(\eta)$	$T(x)$	$\hat{\eta}_k$, for $k \in \{0,1\}$
Normal (known σ^2)	$A(\eta) = \frac{\eta^2}{2}$	$T(x) = \frac{x}{\sigma}$	$\hat{\eta}_k = \frac{1}{N_k\,\sigma} \sum_{i:Y_i=k} X_i$
Normal	$A(\eta) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\log(-2\eta_2)$	$T(x) = \begin{pmatrix} x \\ x^2 \end{pmatrix}$	$\hat{\eta}_k = \begin{pmatrix} \hat{\mu}_k/\hat{\sigma}_k^2 \\ -\frac{1}{2\hat{\sigma}_k^2} \end{pmatrix}, \hat{\mu}_k = \frac{1}{N_k} \sum_{i:Y_i=k} x_i,\ \hat{\sigma}_k^2 = \frac{1}{N_k} \sum_{i:Y_i=k} (x_i - \hat{\mu}_k)^2$
Laplace (known μ)	$A(\eta) = \log(-\frac{2}{\eta})$	$T(x) = \lvert x - \mu \rvert$	$\hat{\eta}_k = -\frac{N_k}{\sum_{i:Y_i=k} \lvert X_i - \mu \rvert}$
Exponential	$A(\eta) = -\log(-\eta)$	$T(x) = x$	$\hat{\eta}_k = -\frac{N_k}{\sum_{i:Y_i=k} X_i}$
Weibull (known k')	$A(\eta) = \log(-\frac{1}{\eta k'})$	$T(x) = x^{k'}$	$\hat{\eta}_k = -\frac{N_k}{\sum_{i:Y_i=k} X_i^{k'}}$

Table 1: We present for five common exponential family distributions their exponential family parametrization functions $A(\eta)$ and $T(x)$ along with their MLEs $\hat{\eta}_k$ for $k \in \{0,1\}$ under EFDA assumptions. Note that for the Normal Distribution (second row), the natural parameter η is a two-dimensional vector.

5 Related Work

While Linear & Quadratic Discriminant Analysis techniques are a well-studied area of research, research on generalizing LDA-style procedures for generative classification is much more sparse. Specifically, to the best of our knowledge, we are the first to explicitly generalize LDA to *generative classification* settings in which class-conditional densities fall under the exponential family.

With that in mind, we present a brief overview of previous related research in this area. To start, [6] presents a procedure they coin "Generalized Discriminant Analysis" to aid in dimensionality reduction by transforming the original feature space through some kernel function. Note however that their application is neither for generative classification nor generalized to all exponential family distributions. Similarly, [3] presents a generalized discriminant analysis procedure through kernel exponential families again for the purpose of dimensionality reduction.

6 Conclusion

We have presented EFDA, a principled extension of LDA to the exponential-family setting. By leveraging the sufficient-statistic representation, EFDA retains LDA's interpretability and closed-form estimators while accommodating a diverse array of distributions. Through simulation with Weibull data, we demonstrated that EFDA can accurately approximate nonlinear log-odds functions, which is beyond the reach of standard logistic regression and LDA. Furthermore, our derivations for four additional distributions in the exponential family underscore EFDA's broad applicability.

We suggest the following few directions for future research:

- **Regularization:** Incorporating a regularization constant to the likelihood function to minimize overfitting in settings where parameters are high dimensional.
- **Exponential Family Mixtures:** Extending EFDA to cases where class–conditional distributions are mixtures within the exponential family.
- **Efficiency as $n \to \infty$:** Comparing the asymptotic variance of EFDA's natural parameter MLEs to that of Logistic Regression's iteratively-solved parameter MLEs.

Our hope with this work is to promote renewed attention at fundamental closed-form & interpretable classification techniques apt for modern non-Gaussian data scenarios.

References

[1] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[2] Trevor Hastie. The elements of statistical learning: data mining, inference, and prediction, 2009.

[3] Isaías Ibañez, Liliana Forzani, and Diego Tomassi. Generalized discriminant analysis via kernel exponential families. *Pattern Recognition*, 132:108933, 2022.

[4] Geoffrey J McLachlan. *Discriminant analysis and statistical pattern recognition*. John Wiley & Sons, 2005.

[5] Frank Nielsen and Vincent Garcia. Statistical exponential families: A digest with flash cards. `https://arxiv.org/abs/0911.4863`, 2009. arXiv:0911.4863.

[6] Shailendra Singh and Sanjay Silakari. Generalized discriminant analysis algorithm for feature reduction in cyber attack detection system. *arXiv preprint arXiv:0911.0787*, 2009.

7 Appendix

7.1 Derivation of EFDA MLEs

For the derivation of $\hat{\alpha}$:

$$0 = \frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{n} \frac{\mathbf{1}\{Y_i = 1\}}{\alpha} - \frac{\mathbf{1}\{Y_i = 0\}}{1-\alpha} \implies (1-\alpha)\sum_{i=1}^{n} \mathbf{1}\{Y_i = 1\} = \alpha \sum_{i=1}^{n} \mathbf{1}\{Y_i = 0\} \implies \hat{\alpha} = \frac{N_1}{n}$$

For the derivation of $\hat{\eta}_1$:

$$0 = \frac{\partial \mathcal{L}}{\partial \eta_1} = \sum_{i=1}^{n} \mathbf{1}\{Y_i = 1\}[T(X_i) - \frac{\partial A(\eta_1)}{\partial \eta_1}] \implies \sum_{i:Y_i=1}^{n} T(X_i) = N_1 \cdot \frac{\partial A(\eta_1)}{\partial \eta_1}$$

and by identical logic our MLE for η_0 must satisfy:

$$\sum_{i:Y_i=0}^{n} T(X_i) = N_0 \cdot \frac{\partial A(\eta_0)}{\partial \eta_0}$$

7.2 Verification of Weibull Exponential Family Parametrization

Using the given exponential family parametrization, we have:

$$h(x)\exp[\eta \cdot T(x) - A(\eta)] = x^{k-1}\exp[-\frac{x^k}{\lambda^k} + \log(\frac{k}{\lambda^k})] = x^{k-1} \cdot \frac{k}{\lambda^k}\exp[-(\frac{x}{\lambda})^k] = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}\exp[-(\frac{x}{\lambda})^k]$$

which is exactly the Weibull(λ, k) PDF.