A. Supplementary Material - Proof of Theorem 3

For convenience we restate the theorem.

Theorem 3. *Define $\mathcal{H}_0$ and $\mathcal{H}_1$ as in (1). For an algorithm with $\beta_n = \mathcal{O}(\frac{1}{n})$, a slack $\gamma = \frac{12np\beta_{n(1-p)}+2\sqrt{2\log(4/\delta)/np}+2\eta}{(1-p)(1+2p)}$, where $p = \frac{|J|}{n}$, and any $\delta \in (0,1)$, we have that under $\mathcal{H}_0$, with probability of at least $1-\delta$,*

$$\left|\hat{R}_{ord} - \hat{R}_{perm}\right| \leq 6np\beta_{n(1-p)} + \sqrt{2\log(4/\delta)/np} + \Delta(1-p)(1+2p) + \eta.$$

Under $\mathcal{H}_1$, the inverse inequality holds with probability of at least $1-\delta$.

Examples for stable algorithms with $\beta_n = \mathcal{O}(\frac{1}{n})$ are k-Nearest Neighbors, soft margin SVM, SVM regression (SVR), and Regularized Least Squares (Bousquet & Elisseeff, 2002).

The proof of the theorem requires the following definitions and lemmas.

We would like to bound the variability of the risk resulting from different training sets. We use the notion of error stability (Kearns & Ron, 1999):

Definition 5. *An algorithm A has error stability β_n with respect to the loss function ℓ if*

$$\forall Z_n \in \mathcal{Z}^n, \forall i \in \{1,...,n\} \ \left|R_{\mathcal{D}}(A_{Z_n}) - R_{\mathcal{D}}(A_{Z_n^{-i}})\right| \leq \beta_n,$$

where Z_n^{-i} is the set Z_n with the sample i removed, and β_n decreases with n.

We extend Definition 5 for the case of $q_n = q(n)$ samples removed.

Definition 6. *An algorithm A has q_n-error stability $\tilde{\beta}_{q_n}$ with respect to the loss function ℓ if*

$$\forall Z_n \in \mathcal{Z}^n, \forall I \subset \{1,...,n\}, |I| = q_n \ \left|R_{\mathcal{D}}(A_{Z_n}) - R_{\mathcal{D}}(A_{Z_n^{-I}})\right| \leq \tilde{\beta}_{q_n},$$

where Z_n^{-I} is the set Z_n with the sample set I removed, and $\tilde{\beta}_{q_n}$ decreases as a function of n.

The triangle inequality implies the following lemma.

Lemma 7. *If algorithm A has error stability β_n and $q_n\beta_{n-q_n}$ decreases as a function of n, then A has q_n-error stability with $\tilde{\beta}_{q_n} \leq q_n\beta_{n-q_n}$.*

Proof. Let $I = \{i_1,...,i_q\}$ be the set of indices removed from Z_n, then if β_n is non-increasing the following holds by the triangle inequality.

$$\left|R_{\mathcal{D}}(A_{Z_n}) - R_{\mathcal{D}}(A_{Z_n^{-I}})\right| \leq \left|R_{\mathcal{D}}(A_{Z_n}) - R_{\mathcal{D}}(A_{Z_n^{-i_1}})\right| + \left|R_{\mathcal{D}}(A_{Z_n^{-i_1}}) - R_{\mathcal{D}}(A_{Z_n^{-I}})\right|$$

$$\beta_n + \left|R_{\mathcal{D}}(A_{Z_n^{-i_1}}) - R_{\mathcal{D}}(A_{Z_n^{-I}})\right| \leq \sum_{i=1}^{q_n} \beta_{n-q_n+i} \leq q_n\beta_{n-q_n+1} \leq q_n\beta_{n-q_n}.$$

Therefore, $\tilde{\beta}_{q_n} \leq q_n\beta_{n-q_n}$. If also $q_n\beta_{n-q_n}$ decreases as a function of n, than $\tilde{\beta}_{q_n}$ decreases with n. □

Corollary 8. *Let algorithm A be with error stability $\beta_n = \mathcal{O}(\frac{1}{n})$, then for $q_n = np$, where $p = \mathcal{O}(\frac{1}{n^c})$, $c \in (0,1]$, algorithm A is also pn-error stable.*

Proof. We have that $np\beta_{n(1-p)} = \mathcal{O}(\frac{np}{n(1-p)}) = \mathcal{O}(\frac{1}{n^c-1})$, which decreases as a function of n. Therefore, by Lemma 7, A also has pn-error stability with $\tilde{\beta}_{pn} \leq np\beta_{n(1-p)}$. □

The following lemma gives a concentration result of the empirical risk over samples $z_t \in S'$, each generated by a distribution $\mathcal{D}_t$.

Lemma 9. *Let S and S' be some train-test split of Z_n, and denote I'_S the index set of S', then for any $\epsilon \in (0,1)$:*

$$\mathbb{P}\left\{\left|R_{I_{S'}}(A_S) - \hat{R}_{S'}(A_S)\right| \leq \epsilon\right\} \geq 1 - 2e^{-2np\epsilon^2}.$$

Proof. By definition

$$\begin{aligned}\mathbb{P}\left\{\left|R_{I_{S'}}(A_S) - \hat{R}_{S'}(A_S)\right| \geq \epsilon\right\} &= \mathbb{P}\left\{|\frac{1}{np}\sum_{i\in I_{S'}} R_{\mathcal{D}_i}(A_S) - \ell(A_S, z_i)| \geq \epsilon\right\} \\ &= \mathbb{E}_S\left[\mathbb{P}_{S'}\left[|\frac{1}{np}\sum_{i\in I_{S'}} R_{\mathcal{D}_i}(A_S) - \ell(A_S, z_i)| \geq \epsilon \middle| S\right]\right].\end{aligned}$$

Next, we bound the inner conditional probability using Hoeffding's inequality: $\mathbb{E}\left[e^{\alpha X}\right] \leq e^{\alpha^2/8}$ for a zero mean random variable in the range $0 \leq X \leq 1$ and $\alpha > 0$.

$$\begin{aligned}&\mathbb{P}_{S'}\left[\frac{1}{np}\sum_{i\in I_{S'}}(\ell(A_S, z_i) - R_{\mathcal{D}_i}(A_S)) \geq \epsilon \middle| S\right] \leq \inf_{\lambda>0} e^{-\lambda\epsilon}\mathbb{E}_{S'}\left[e^{\frac{\lambda}{np}\sum_{i\in I_{S'}}\left(\ell(A_S,z_i) - R_{\mathcal{D}_i}(A_S)\right)} \middle| S\right] \\ &\leq \inf_{\lambda>0} e^{-\lambda\epsilon}\prod_{i\in I_{S'}}\mathbb{E}_{S'}\left[e^{\frac{\lambda}{np}\left(\ell(A_S,z_i) - R_{\mathcal{D}_i}(A_S)\right)} \middle| S\right] \leq \inf_{\lambda>0} e^{-\lambda\epsilon}e^{\frac{\lambda^2}{8np}} = e^{-2np\epsilon^2}.\end{aligned}$$

The complimentary bound can be derived similarly. □

The following components d_0 and d_1 differentiate between the behavior under the null and alternative hypothesis. The lemma gives upper and lower bounds on these components.

Lemma 10. *Let I and J be two sets of consecutive indices ranges, such that $|J| = \lceil pn \rceil$ and $|I| = n - |J|$ for some $p \in (0,1)$, and $IJ = I \cup J$. Let A be a pn-error stable algorithm (Definition 6). Denote the difference between the expected risks of J and IJ of the function f by $d_0 \doteq R_J(f) - R_{IJ}(f)$, and the difference between the expected risks over all (S, S') train-test splits by $d_1 \doteq \mathbb{E}_{S\sim\mathcal{U}_n}\left[R_{IJ}(A_S)\right] - \mathbb{E}_{S\sim\mathcal{U}_n}\left[R_{I_{S'}}(A_S)\right]$.*

Under $\mathcal{H}_0 : |R_J(f) - R_I(f)| \leq \Delta$ we have that

$$d_0 \leq (1-p)\Delta,\ d_1 \leq 2p(1-p)\Delta + \eta + 4\tilde{\beta}_{np},$$

Under $\mathcal{H}_1 : |R_J(f) - R_I(f)| \geq \Delta + \gamma$ we have that

$$d_0 \geq (1-p)(\Delta+\gamma),\ d_1 \geq 2p(1-p)(\Delta+\gamma) - \eta - 4\tilde{\beta}_{np}.$$

Proof. For any function f for which $|R_I(f) - R_J(f)| \leq \Delta$ we have that the difference between the average risks is bounded:

$$\begin{aligned}|R_{IJ}(f) - R_I(f)| &= |(1-p)R_I(f) + pR_J(f) - R_I(f)| = p\,|R_J(f) - R_I(f)| \leq p\Delta, \\ |R_{IJ}(f) - R_J(f)| &= |(1-p)R_I(f) + pR_J(f) - R_J(f)| = (1-p)\,|R_J(f) - R_I(f)| \leq (1-p)\Delta.\end{aligned} \tag{3}$$

Therefore, under $\mathcal{H}_0$ we have that $d_0 \doteq R_J(f) - R_{IJ}(f) \leq (1-p)\Delta$.

Due to the η-permitted variation assumption, the difference between the average risk on IJ and the risk with respect to some $\mathcal{D}_i$ for $i \in IJ$ is also bounded:

$$\begin{aligned}&\forall i \in I \quad |R_{IJ}(f) - R_{\mathcal{D}_i}(f)| \leq p\Delta + |R_I(f) - R_{\mathcal{D}_i}(f)| \leq p\Delta + \eta, \\ &\forall j \in J \quad \left|R_{IJ}(f) - R_{\mathcal{D}_j}(f)\right| \leq (1-p)\Delta + \left|R_J(f) - R_{\mathcal{D}_j}(f)\right| \leq (1-p)\Delta + \eta.\end{aligned} \tag{4}$$

By stability of algorithm A we have that for any distribution P, Q :

$$|R_P(A_S) - R_Q(A_S)| \leq |R_P(A_S) - R_P(A_{Z_n})| + |R_Q(A_{Z_n}) - R_Q(A_S)| + |R_P(A_{Z_n}) - R_Q(A_{Z_n})| \leq \quad (5)$$
$$2\tilde{\beta}_{np} + |R_P(A_{Z_n}) - R_Q(A_{Z_n})| \leq 4\tilde{\beta}_{np} + |R_P(f) - R_Q(f)|\,.$$

Recall that $\mathcal{U}_n$ denotes the uniform distribution over all possible (S, S') splits. Notice that the sizes of the sets (S, S') are $|S| = |I|$ and $|S'| = |J|$. Under the null hypothesis and inequalities (4) and (5) we have that

$$|R_{IJ}(A_S) - R_{\mathcal{D}_i}(A_S)| \leq 4\tilde{\beta}_{np} + |R_{IJ}(f) - R_{\mathcal{D}_i}(f)| \leq 4\tilde{\beta}_{np} + p\Delta + \eta, \forall i \in I. \quad (6)$$
$$|R_{IJ}(A_S) - R_{\mathcal{D}_i}(A_S)| \leq 4\tilde{\beta}_{np} + (1-p)\Delta + \eta, \forall j \in J.$$

By inserting the above inequalities we bound the expected difference between the risk on IJ and the risk on the set $I_{S'}$ corresponding to the test set S':

$$d_1 \doteq \mathbb{E}_{S\sim\mathcal{U}_n}\left[R_{IJ}(A_S) - R_{I_{S'}}(A_S)\right] = \frac{1}{np\binom{n}{np}} \sum_{S'} \sum_{i\in S'} (R_{IJ}(A_S) - R_{\mathcal{D}_i}(A_S))$$
$$\leq \frac{1}{np\binom{n}{np}} \binom{n-1}{np-1} \left[(1-p)n(p\Delta + \eta + 4\tilde{\beta}_{np}) + pn((1-p)\Delta + \eta + 4\tilde{\beta}_{np})\right] = 2p(1-p)\Delta + \eta + 4\tilde{\beta}_{np}.$$

The assertions under the alternative hypothesis are obtained in a similar manner to the above derivation. For any function f for which $|R_I(f) - R_J(f)| \geq \Delta + \gamma$ it may be shown that

$$|R_{IJ}(f) - R_I(f)| \geq p(\Delta + \gamma),$$
$$|R_{IJ}(f) - R_J(f)| \geq (1-p)(\Delta + \gamma),$$
$$\forall i \in I \quad |R_{IJ}(f) - R_{\mathcal{D}_i}(f)| \geq p(\Delta + \gamma) - \eta,$$
$$\forall j \in J \quad \left|R_{IJ}(f) - R_{\mathcal{D}_j}(f)\right| \geq (1-p)(\Delta + \gamma) - \eta.$$

Combined with the stability of the algorithm we have that

$$d_0 \geq (1-p)(\Delta + \gamma) \text{ and } d_1 \geq 2p(1-p)(\Delta + \gamma) - \eta - 4\tilde{\beta}_{np}. \quad (7)$$

□

We are now ready to provide the proof of the main theorem.

Proof (Theorem 3): We begin by showing that the following holds:

$$\mathbb{P}\left\{\left|\hat{R}_{ord} - \hat{R}_{perm} - d_0 - d_1\right| \geq 2np\beta_{n(1-p)} + \sqrt{2\log(4/\delta)/np}\right\} \leq \delta, \quad (8)$$

where $d_0 \doteq R_J(f) - R_{IJ}(f)$, and $d_1 \doteq \mathbb{E}_{S\sim\mathcal{U}_n}\left[R_{IJ}(A_S)\right] - \mathbb{E}_{S\sim\mathcal{U}_n}\left[R_{I_{S'}}(A_S)\right]$ are as defined in Lemma 10.
The bound is obtained as follows:

$$\left|\hat{R}_{ord} - \hat{R}_{perm} - d_0 - d_1\right| \doteq$$
$$\left|\hat{R}_{ord} - \mathbb{E}_{S\sim\mathcal{U}_n}\left[\hat{R}_{S'}(S)\right] - R_J(f) + R_{IJ}(f) - \mathbb{E}_{S\sim\mathcal{U}_n}\left[R_{IJ}(A_S) - R_{I_{S'}}(A_S)\right]\right| \leq$$
$$\left|\hat{R}_{ord} - R_J(f)\right| + \mathbb{E}_{S\sim\mathcal{U}_n}\left[\left|R_{I_{S'}}(A_S) - \hat{R}_{S'}(A_S)\right|\right] + \left|R_{IJ}(f) - \mathbb{E}_{S\sim\mathcal{U}_n}\left[R_{IJ}(A_S)\right]\right|.$$

The probability of the first two components is bounded by Lemma 9. The probability of the last component is bounded by stability as follows. By construction, $|Z_n| - |S| = |J| = np$. By Corollary 8 algorithm A is pn-error stable with $\tilde{\beta}_{pn} \leq np\beta_{n(1-p)}$. Therefore,

$$|R_{IJ}(f) - \mathbb{E}_{S\sim\mathcal{U}_n}\left[R_{IJ}(A_S)\right]| \leq$$
$$|R_{IJ}(f) - R_{IJ}(A_{Z_n})| + |R_{IJ}(A_{Z_n}) - \mathbb{E}_{S\sim\mathcal{U}_n}\left[R_{IJ}(A_S)\right]| \leq$$
$$|R_{IJ}(f) - R_{IJ}(A_{Z_n})| + \max_{S\sim\mathcal{U}_n} |R_{IJ}(A_{Z_n}) - R_{IJ}(A_S)| \leq 2\tilde{\beta}_{pn} \leq 2np\beta_{n(1-p)}$$

with probability one.

By Lemma 10, we have that under the null hypothesis $d_0 + d_1 \leq \Delta(1-p)(1+2p) + \eta + 4\tilde{\beta}_{np}$. Inserting this bound to the inequality in (8) we have that

$$\mathbb{P}_{\mathcal{H}_0}\left[\hat{R}_{ord} - \hat{R}_{perm} \geq 6np\beta_{n(1-p)} + \sqrt{2\log(4/\delta)/np} + \Delta(1-p)(1+2p) + \eta\right] \leq \delta,$$

which proves the first part of the theorem.

For the inverted inequality, first recall that $\gamma = \frac{12np\beta_{n(1-p)} + 2\sqrt{2\log(4/\delta)/np} + 2\eta}{(1-p)(1+2p)}$.
By Lemma 10

$$d_0 + d_1 \geq (\Delta + \gamma)(1-p)(1+2p) - \eta - 4np\beta_{n(1-p)} = \Delta(1-p)(1+2p) + 8np\beta_{n(1-p)} + 2\sqrt{2\log(4/\delta)/np} + \eta \tag{9}$$

Under the alternative hypothesis, the probability of obtaining a false negative is bounded as follows.

$$\mathbb{P}_{\mathcal{H}_1}\left[\left|\hat{R}_{ord} - \hat{R}_{perm}\right| \leq 6np\beta_{n(1-p)} + \sqrt{2\log(4/\delta)/np} + \Delta(1-p)(1+2p) + \eta\right] =$$

$$\mathbb{P}_{\mathcal{H}_1}\left[\left|\hat{R}_{ord} - \hat{R}_{perm}\right| \leq -2np\beta_{n(1-p)} - \sqrt{2\log(4/\delta)/np} + \left(8np\beta_{n(1-p)} + 2\sqrt{2\log(4/\delta)/np} + \Delta(1-p)(1+2p) + \eta\right)\right] \overset{(a)}{\leq}$$

$$\mathbb{P}_{\mathcal{H}_1}\left[\left|\hat{R}_{ord} - \hat{R}_{perm}\right| \leq -2np\beta_{n(1-p)} - \sqrt{2\log(4/\delta)/np} + d_0 + d_1\right] \leq$$

$$\mathbb{P}_{\mathcal{H}_1}\left[d_0 + d_1 - \hat{R}_{ord} + \hat{R}_{perm} \geq 2np\beta_{n(1-p)} + \sqrt{2\log(4/\delta)/np}\right] \overset{(b)}{\leq} \delta,$$

where (a) holds by Equation (9), and (b) is by applying the bound in (8). □